



**University of  
Zurich**<sup>UZH</sup>

**Zurich Open Repository and  
Archive**

University of Zurich  
University Library  
Strickhofstrasse 39  
CH-8057 Zurich  
[www.zora.uzh.ch](http://www.zora.uzh.ch)

---

Year: 2020

---

## **Tell us about your leadership style: A structured interview approach for assessing leadership behavior constructs**

Heimann, Anna Luca ; Ingold, Pia Verena ; Kleinmann, Martin

**Abstract:** It is widely recognized that leadership behaviors drive leaders' success. But despite the importance of assessing leadership behavior for selection and development, current measurement practices are limited. This study contributes to the literature by examining the structured interview method as a potential approach to assess leadership behavior. To this end, we developed a structured interview measuring constructs from Yukl's (2012) leadership taxonomy. Supervisors in diverse positions participated in the interview as part of a leadership assessment program. Confirmatory factor analyses supported the assumption that leadership constructs could be assessed as distinct interview dimensions. Results further showed that interview ratings predicted a variety of leadership outcomes (supervisors' annual income, ratings of situational leader effectiveness, subordinates' well-being and affective organizational commitment) beyond other relevant predictors. Findings offer implications on how to identify leaders who have a positive impact on their subordinates, and they inform us about conceptual differences between leadership measures.

DOI: <https://doi.org/10.1016/j.leaqua.2019.101364>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-191280>

Journal Article

Published Version



The following work is licensed under a Creative Commons: Attribution 4.0 International (CC BY 4.0) License.

Originally published at:

Heimann, Anna Luca; Ingold, Pia Verena; Kleinmann, Martin (2020). Tell us about your leadership style: A structured interview approach for assessing leadership behavior constructs. *Leadership Quarterly*, 31(4):101364.

DOI: <https://doi.org/10.1016/j.leaqua.2019.101364>



## Full Length Article

Tell us about your leadership style: A structured interview approach for assessing leadership behavior constructs<sup>☆</sup>Anna Luca Heimann<sup>\*</sup>, Pia V. Ingold, Martin Kleinmann

Department of Psychology, University of Zurich, Switzerland

## ARTICLE INFO

## Keywords:

Structured interviews  
Leadership behavior  
Incremental validity  
Well-being  
Affective organizational commitment

## ABSTRACT

It is widely recognized that leadership behaviors drive leaders' success. But despite the importance of assessing leadership behavior for selection and development, current measurement practices are limited. This study contributes to the literature by examining the structured interview method as a potential approach to assess leadership behavior. To this end, we developed a structured interview measuring constructs from Yukl's (2012) leadership taxonomy. Supervisors in diverse positions participated in the interview as part of a leadership assessment program. Confirmatory factor analyses supported the assumption that leadership constructs could be assessed as distinct interview dimensions. Results further showed that interview ratings predicted a variety of leadership outcomes (supervisors' annual income, ratings of situational leader effectiveness, subordinates' well-being and affective organizational commitment) beyond other relevant predictors. Findings offer implications on how to identify leaders who have a positive impact on their subordinates, and they inform us about conceptual differences between leadership measures.

Leadership behaviors have been shown to relate to indicators of organizational performance such as leader effectiveness and subordinate outcomes (Harms, Credé, Tynan, Leon, & Jeung, 2017; Jackson, Meyer, & Wang, 2013; Judge & Piccolo, 2004). Given the impact of leadership behaviors on organizational outcomes, a careful assessment of these behaviors is a relevant step in selecting and developing supervisors.

Leadership behavior constructs are usually assessed via questionnaires that are filled in either by subordinates or by supervisors themselves (Hunter, Bedell-Avers, & Mumford, 2007). However, research has shown that ratings from both sources can be substantially biased (e.g., Fleenor, Smither, Atwater, Braddy, & Sturm, 2010; Hansbrough, Lord, & Schyns, 2015). Moreover, leadership questionnaires tend to be generic and do not take into account the situation in which leadership behaviors occur (Yukl, 1999). This is problematic, given that the effectiveness of leadership behaviors can depend on the specific situation (Liden & Antonakis, 2009; Vroom & Jago, 2007; Zaccaro, Green, Dubrow, & Kolze, 2018).

In light of this problem, there have been strong calls to explore alternative approaches to measuring leadership behavior constructs (Antonakis, Bastardo, Jacquart, & Shamir, 2016; DeRue, Nahrgang,

Wellman, & Humphrey, 2011; Hunter et al., 2007; Yukl, 1999). The structured interview method may be considered a particularly useful alternative to conventional leadership questionnaires, given that structured interviews are established and feasible assessment instruments that do not rely on potentially biased self-ratings or subordinate ratings (Huffcutt, Conway, Roth, & Stone, 2001; Huffcutt, Weekley, Wiesner, DeGroot, & Jones, 2001; Krajewski, Goffin, McCarthy, Rothstein, & Johnston, 2006). Furthermore, structured interviews ask for interviewees' behavior in specific situations, and therefore offer the opportunity for a more context-sensitive assessment of leadership (similar to situational judgment tests; see also Liden & Antonakis, 2009; Peus, Braun, & Frey, 2013). Despite these advantages, research has yet to employ interview methodology for assessing established leadership behavior constructs.

The present study is the first to connect structured interviews with leadership research by assessing leadership behavior constructs as dimensions in a structured interview, and therefore offers several contributions. First, this study investigates the potential of a novel approach to assess leadership behavior constructs that is based on an independent rating source (i.e., interviewer ratings), and that considers the situation in which leadership is exhibited. Second, we address the

<sup>☆</sup> The study reported in this paper was supported by grants from the Swiss National Science Foundation (grant numbers 146039 and 175827).

We thank Gary Yukl and Rolf van Dick for their helpful comments on an earlier version of the manuscript. We also thank Dolores Arias, Alessandro Baia, Sabine Benz, Debora Exer, Deborah Lüdin, Magdalena Mehl, and Michael Uehlinger, who helped collect the data for this study.

<sup>\*</sup> Corresponding author at: Work and Organisational Psychology, University of Zurich, Binzmühlestrasse 14/12, 8050 Zurich, Switzerland.

E-mail address: [a.heimann@psychologie.uzh.ch](mailto:a.heimann@psychologie.uzh.ch) (A.L. Heimann).

antagonistic relationship between effectiveness and well-being in organizational research (Kozlowski, 2012) by incorporating both outcomes regarding leaders' performance and outcomes regarding subordinates' well-being when examining the interviews' validity. Third, to deepen our understanding of the interview method, this study explores how interview ratings of leadership behavior constructs correspond to a variety of different approaches to assess leadership (i.e., self-ratings, subordinate ratings, and behavioral codings).

### Assessing leadership behaviors

While the leadership literature has produced a broad range of valuable constructs to describe leadership behaviors (e.g., Antonakis & House, 2014; Avolio, Bass, & Jung, 1999; Conger & Kanungo, 1998; Stogdill & Coons, 1957), there has been a growing demand for more conceptual clarity in leadership frameworks. For instance, various authors have argued that there is substantial conceptual and empirical overlap between different leadership behavior constructs both within and across frameworks (e.g., Piccolo et al., 2012; Shaffer, DeGeest, & Li, 2016; Yukl, 1999).

To address this lack of conceptual clarity, Yukl, Gordon, and Taber (2002) clustered different leadership behaviors from existing frameworks into three broad meta-categories: task-oriented leadership (e.g., explaining tasks and responsibilities, planning and prioritizing activities), relations-oriented leadership (e.g., providing individual support and encouragement, recognizing achievements), and change-oriented leadership (e.g., communicating a vision of what can be accomplished, explaining why changes are needed). This taxonomy only includes leadership behaviors from existing frameworks that are (a) directly observable, (b) potentially relevant to all types of supervisors within and across organizations, and (c) clearly assigned to only one category of leadership behaviors (Yukl, 2012; Yukl et al., 2002).

Although Yukl's taxonomy provides a common ground to conceptualize leadership constructs, a remaining question is *how* to obtain meaningful ratings of these constructs in the context of selection and development. Self-ratings of leadership behaviors, for example, may need to be interpreted with caution, given that a number of factors can produce systematic bias in the perception of one's own typical behavior (i.e., personality and demographic characteristics; see Fleenor et al., 2010; Gentry, Hannum, Ekelund, & de Jong, 2007; Sala, 2003). Accordingly, previous research suggests that self-ratings do not reflect actual leadership behaviors, but rather they provide insight into the characteristics of the self-rater (e.g., supervisors' self-esteem and self-awareness; Atwater & Yammarino, 1992; Goffin & Anderson, 2007).

Similarly, subordinate ratings of leadership behaviors are prone to different types of biases (for an overview see Fleenor et al., 2010). For example, subordinates tend to evaluate their supervisors' everyday behavior more favorably when supervisors' attributes correspond to subordinates' implicit assumptions about what a supervisor should be like (i.e., leader prototypicality; see Junker & van Dick, 2014) and when they personally like their supervisor (e.g., Rowold & Borgmann, 2014). This indicates that subordinate ratings "may have little to do with actual leader behavior" (Hansbrough et al., 2015, p. 221), but rather reflect relationships of supervisors with their subordinates.

### Structured interviews for assessing leadership behavior

Given the issues involved in current leadership measurement practices, there is a need to explore the potential of alternative approaches to assessing leadership behaviors (e.g., DeRue et al., 2011; Hunter et al., 2007). Although there exists a variety of assessment methods (i.e., structured interviews, situational judgment tests, business games), these instruments have rarely been used for assessing constructs from leadership research (for an exception see Peus et al., 2013). Among these instruments, the structured interview method possesses particular potential as an alternative to conventional leadership questionnaires for

several reasons.

To begin with, structured interviews usually rely on independent rating sources, namely the interviewers. As compared to supervisors or subordinates, interviewers should have less motivation to provide socially desirable ratings because they are usually not involved with the rated supervisor at a personal level. Furthermore, interviewers are often trained on how to avoid biases and on how to systematically evaluate the constructs that are to be assessed (e.g., Roch, Woehr, Mishra, & Kieszczyńska, 2012).

In addition, structured interview questions allow for a more situation-specific assessment of leadership behaviors as compared to conventional leadership questionnaires. This is because the most common structured interview question types typically ask how interviewees behaved or would behave in actually experienced situations in the past (i.e., past-oriented interview questions; Janz, 1982) or in hypothetical situations (i.e., future-oriented interview questions; Latham, Saari, Pursell, & Campion, 1980). Thus, structured interview questions provide interviewees with a more detailed context as compared to conventional questionnaires that consist of generic items and do not present context information (e.g., "My supervisor leads by example" or "My supervisor will not settle for second best"; Podsakoff, MacKenzie, Moorman, & Fetter, 1990).

Given their high level of contextualization, structured interviews may function in a similar way as other situation-specific assessment tools such as situational judgment tests. The underlying rationale of situation-specific assessments is that the assessed person is first required to identify the demands of a given situation, and then has to present an adequate reaction to this situation (Jansen et al., 2013; Kleinmann et al., 2011). Hence, structured interviews capture (a) interviewees' cognitive understanding of the presented (past or future) situations (see also Fan, Stuhlmán, Chen, & Weng, 2016; Melchers & Kleinmann, 2016), and (b) interviewees' past experiences, specific knowledge, and general beliefs about which behaviors are effective in these situations (Lievens & Motowidlo, 2016; Motowidlo & Beier, 2010).

Despite these advantages of the interview methodology, only few studies have examined structured interviews specifically for assessing leadership behavior (Huffcutt, Weekley, et al., 2001; Krajewski et al., 2006). Huffcutt, Weekley, et al. (2001) found that interview questions that are designed to assess various leadership behaviors have the potential to predict supervisors' performance as rated by superiors (i.e., supporting the interviews' criterion-related validity). However, they did not find evidence supporting that the internal data structure of interview ratings represented the intended interview dimensions. In particular, correlations were low (i.e.,  $r = 0.09$  in Sample 1 and  $r = 0.05$  in Sample 2) between past-oriented and future-oriented interview questions that were designed to assess the same leadership behaviors. This raises questions as to whether interview questions adequately assessed the leadership behaviors that they intended to measure. Krajewski et al. (2006) replicated this study; they also used superiors' ratings as the sole criterion measure and found results similar to those from Huffcutt, Weekley, et al. (2001).

In sum, the main limitations of previous research are (a) that interviews were not designed to assess constructs from existing leadership frameworks, (b) that there is little evidence that existing leadership interviews measure the leadership behaviors that they intend to measure, and (c) that only ratings from supervisors' superiors have been used as criterion measure, even though leadership behaviors may primarily affect subordinates and, thus, subordinate outcomes have been seen as important criterion measures in leadership research (Hiller, DeChurch, Murase, & Doty, 2011).

### The present study

Drawing from a comprehensive leadership taxonomy (Yukl, 2012; Yukl et al., 2002), the present study explores the potential of the

structured interview method to assess leadership behavior constructs. To this end, it examines whether the internal data structure of interview ratings reflects the leadership behavior constructs that the interview is designed to measure; whether interview ratings can predict different types of leadership outcomes over and above other relevant predictors; and to what extent interview ratings relate to other leadership measures.

#### *Internal structure of interview ratings*

Generally, previous research has found little evidence that the internal data structure of interview ratings represents the constructs or interview dimensions that the structured interviews were actually designed to measure (Huffcutt, Weekley, et al., 2001; Macan, 2009; Van Iddekinge, Raymark, Eidson Jr., & Attenweiler, 2004). Researchers often find that the interviews' data structure is best represented by factor models that do not specify different interview dimensions in confirmatory factor analyses (CFAs; see for example Klehe, König, Richter, Kleinmann, & Melchers, 2008; Krajewski et al., 2006). A reason for this consistent finding might be that structured interviews often do not assess well-defined leadership constructs but unspecific and ad-hoc labeled interview dimensions such as "leadership behaviors" (Melchers et al., 2009).

Despite these findings, we posit that it is possible to assess task-, relations-, and change-oriented leadership as separate interview dimensions for two reasons: First, they are conceptually distinct and clearly defined constructs indicated by specific, observable behaviors (Yukl, 2012; Yukl et al., 2002). Second, there is initial empirical evidence that these leadership behavior constructs can be measured as different dimensions: Previous studies conducting CFAs found that factor models including task-, relations-, and change-oriented leadership as separate factors fitted the data structure of questionnaire-based leadership ratings (Borgmann, Rowold, & Bormann, 2016; Yukl et al., 2002). Accordingly, we hypothesize:

**Hypothesis 1.** A factor model specifying task-, relations-, and change-oriented leadership as distinct factors will best represent the internal data structure of interview ratings.

#### *Interview ratings and leadership outcomes*

The utility of the structured interview approach for leader selection and development ultimately depends on whether interview ratings can predict relevant leadership outcomes (i.e., show criterion-related validity). When studying outcomes of leadership, it has been recommended to "include a variety of criteria" (Yukl, 2013, p. 26). In their integrative meta-analytic review of the leadership literature, DeRue et al. (2011) propose to distinguish between three content dimensions of leadership outcomes: (a) the overall judgment of leaders' success; (b) performance-focused outcomes; and (c) affective outcomes. To cover such a proposed variety of leadership outcomes, this study examines variables indicative of each category of outcomes: supervisors' reported annual income as an indicator of overall success; ratings of supervisors' situational effectiveness in behavioral leadership tasks as a performance-focused outcome; and ratings of subordinates' general well-being and affective organizational commitment as affective outcomes. We included two affective outcomes to cover different foci: Subordinates' general well-being helps studying the overall impact of supervisors on their subordinates' health and psychological functioning without referring to the context of work (e.g., Montano, Reeske, Franke, & Hüffmeier, 2017), while subordinates' affective organizational commitment is a more specific outcome telling us to what extent supervisors shape subordinates' attitude and attachment to their specific job (e.g., Jackson et al., 2013).

In determining the utility of the interview method, the final question is whether interview ratings explain incremental variance in

leadership outcomes over and above other leadership predictors. Given that incremental validity evidence needs to be grounded in theory (Antonakis & Dietz, 2011), we drew from DeRue et al.'s (2011) integrative framework that distinguishes between two common classes of predictors of effective leadership: Leader traits and leadership behaviors.

Leader traits are stable person characteristics that can be grouped into three categories (DeRue et al., 2011): (a) characteristics that refer to demographics, (b) characteristics associated with task competence, and (c) characteristics describing interpersonal competencies. The present study includes predictors from each of these categories: age, gender, and leadership experience as demographic characteristics; supervisors' core self-evaluations as a competence-related leader trait; and supervisors' emotional intelligence as an interpersonal leader trait. All these leader characteristics have been shown to be related to leadership outcomes (Ahn, Lee, & Yun, 2018; Hoffman, Woehr, Maldagen-Youngjohn, & Lyons, 2011; Judge & Kammeyer-Mueller, 2011; Miao, Humphrey, & Qian, 2016; Paustian-Underdahl, Walker, & Woehr, 2014; Walter & Scheibe, 2013). Yet, we expect interview ratings of leadership behaviors to outperform these traits given that behaviors are considered more proximal predictors to leadership outcomes than traits (DeRue et al., 2011).

Leadership behavior constructs describe how supervisors act towards their subordinates (Yukl et al., 2002). Given the feasibility of assessing leadership behavior constructs by asking supervisors or subordinates to fill out a leadership questionnaire (e.g., Hunter et al., 2007), it seems relevant to examine whether interview ratings are stronger predictors of leadership outcomes than questionnaires when both types of measures were designed to assess the same leadership constructs. As discussed earlier, we expect interview ratings to be more predictive of leadership outcomes than questionnaire-based self-ratings and subordinate ratings of leadership behaviors, given that interview ratings rely on independent and trained rating sources (i.e., interviewers; see also Roch et al., 2012) and allow for a more context-specific assessment of leadership behaviors (similar to situational judgment tests; Peus et al., 2013). Taken together, we explore the incremental criterion-related validity of the interview method by examining whether the overall interview rating score of leadership behaviors predict different leadership outcomes over and above leader traits and also self-ratings and subordinate-ratings of leadership behavior.

#### *Interview ratings and supervisors' annual income*

Generally, supervisors engage in leadership behaviors with the objective of being successful at their jobs. Thereby, they aim to advance their organization, as well as their own career and income. In support of this assumption, previous research has demonstrated (a) that higher scores on leadership behavior constructs relate to higher levels of individual and organizational success (e.g., Ceri-Booms, Curşeu, & Oerlemans, 2017; Geyer & Steyrer, 1998; Wilderom, van den Berg, & Wiersma, 2012), and (b) that self-reported annual income can be regarded as an indicator of individual success (e.g., Judge, Klinger, & Simon, 2010; Ng, Eby, Sorensen, & Feldman, 2005; Rode, Arthaud-Day, Ramaswami, & Howes, 2017). Accordingly, we expect supervisors who demonstrate higher behavioral leadership competencies in the interview to be more successful at their jobs and therefore to earn greater annual income. Considering that annual income is a relatively broad indicator of success that is potentially confounded with a number of variables (e.g., age, leadership experience, gender, industry, size of the organization), we will control for these confounding variables in examining the validity of the overall interview in predicting annual income:

**Hypothesis 2.** The overall interview score explains a significant level of variance in supervisors' reported annual income over and above the size of the organization, industry, demographic variables, core self-



evaluations, emotional intelligence, and the overall scores of self-rated and subordinate-rated leadership behaviors.

#### *Interview ratings and ratings of situational leader effectiveness*

Leader effectiveness is a frequently studied outcome and describes how well supervisors perform their job (i.e., their leadership tasks; see Hiller et al., 2011). Per definition, actively engaging in leadership behaviors should help supervisors to master their leadership tasks (Yukl, 2012). In support of this, leadership behavior constructs (measured via leadership questionnaires) have often been found to meta-analytically relate to leader effectiveness (e.g., DeRue et al., 2011; Judge & Piccolo, 2004).

The present study aims to collect ratings of leader effectiveness from a rating source that can evaluate supervisors' performance more independently than potentially biased subordinates (e.g., Rowold & Borgmann, 2014; Wang, Van Iddekinge, Zhang, & Bishoff, 2019). Specifically, we collect ratings from role-players who evaluate supervisors' situational effectiveness in standardized behavioral leadership tasks (i.e., in simulated situations). In this context, role-players' effectiveness ratings serve as an indicator of supervisors' performance in the presented tasks. The advantage of role-players is that they should not have any motive to provide biased ratings (i.e., they have no long-term commitment towards the person they rate). In addition, role-players evaluate the situational effectiveness of all supervisors in the same standardized leadership tasks, thereby increasing the comparability of ratings across supervisors. We posit the following hypothesis:

**Hypothesis 3.** The overall interview score explains a significant level of variance in ratings of situational leader effectiveness over and above demographic variables, core self-evaluations, emotional intelligence, and the overall scores of self-rated and subordinate-rated leadership behaviors.

#### *Interview ratings and subordinates' general well-being*

More recently, subordinate well-being has received considerable attention in the leadership literature (Kurtessis et al., 2017; Montano et al., 2017). It is defined as a state of positive mental health in which an individual often feels cheerful, active, fresh, and rested (Topp, Østergaard, Søndergaard, & Bech, 2015; WHO, 2001). Previous research suggests that supervisors whose leadership behaviors support their subordinates at work (e.g., by providing sufficient resources to accomplish work objectives, recognizing accomplishments, explaining why subordinate's work is relevant; Yukl, 2012) can serve as a work resource for subordinates and thereby foster subordinate well-being (see Nielsen et al., 2017). Accordingly, a meta-analysis showed that leadership behavior constructs (measured via questionnaires) significantly relate to subordinates' well-being (Montano et al., 2017). Hence, we posit that supervisors who are rated higher on leadership behavior constructs in the interview will have subordinates with higher levels of general well-being:

**Hypothesis 4.** The overall interview score explains a significant level of variance in subordinates' well-being over and above supervisors' demographic variables, core self-evaluations, emotional intelligence, and the overall scores of self-rated and subordinate-rated leadership behaviors.

#### *Interview ratings and subordinates' affective organizational commitment*

Affective organizational commitment refers to subordinates' emotional attachment to or involvement in an organization (Allen & Meyer, 1990), and it is considered an attitudinal outcome of leadership (Borgmann et al., 2016; Jackson et al., 2013; Schyns & Schilling, 2013).

The rationale behind this is that supervisors whose leadership behaviors support their subordinates create favorable working conditions (see also Kurtessis et al., 2017). From a social exchange perspective (see Cropanzano & Mitchell, 2005), supervisors creating a positive work environment should help subordinates to develop favorable attitudes towards their job and organization, including the desire to stay with the organization (Jackson et al., 2013). In support of this, meta-analyses found leadership behavior constructs to relate substantially to subordinates' affective organizational commitment (Borgmann et al., 2016; Jackson et al., 2013). Thus, we expect that supervisors who demonstrate higher behavioral leadership competencies in the interview will have subordinates who report higher levels of affective organizational commitment:

**Hypothesis 5.** The overall interview score explains a significant level of variance in subordinates' affective commitment over and above supervisors' demographic variables, core self-evaluations, emotional intelligence, and the overall scores of self-rated and subordinate-rated leadership behaviors.

#### *Interview ratings and other measures of leadership*

Finally, to deepen our understanding of the interview methodology, we further explore relationships between interview ratings and other measures of leadership (i.e., questionnaire-based self-ratings, subordinate ratings and behavioral codings of leadership). Previous research generally indicates little convergence between interview ratings and other types of measures (Atkins & Wood, 2002; Van Iddekinge et al., 2004; Van Iddekinge, Raymark, & Roth, 2005), which adds weight to the question of what type of interview ratings capture. No study has yet examined the extent to which interview ratings correspond to other measures that were designed to assess the *same* leadership constructs (i.e., to what extent structured interviews are either redundant to or complement other measures). We therefore pose the following research question:

**Research Question.** To what extent do interview ratings of task-, relations- and change-oriented leadership relate to other measures of leadership behavior (i.e., self-ratings, subordinate ratings, or behavioral codings)?

## **Method**

### *Sample*

#### *Supervisors*

The sample consisted of 152 supervisors (62 women, 90 men) who participated in a leadership interview as part of a leadership assessment program. The assessment program was delivered at a center for continuing education in Europe and was open to supervisors from all kinds of organizations. Supervisors participated in the assessment program to receive extensive feedback on their leadership behaviors. Participants were recruited via local print advertising, social media, and by contacting the human resource departments of local organizations. To be included in the present study, supervisors had to allow us to collect data from at least two of their subordinates (a) who they supervised directly, (b) with whom they had been working together for at least six months, and (c) with whom they interacted with several times a week. Supervisors were on average 44.95 ( $SD = 7.41$ ) years old, worked 41.45 ( $SD = 5.06$ ) hours per week, and had 9.32 ( $SD = 6.54$ ) years of leadership experience.

Supervisors worked in different sectors. About 5% of them worked in finance and insurance, 9% in health care and social services, 16% in manufacturing, 5% in media and communication, 6% in non-government organizations, 14% in public services and administration, 20% in research and education, 6% in sales and marketing, 14% as private

service providers, 2% in transport and logistics, and 3% did not indicate any of these categories.

### Subordinates

We collected data from 450 eligible subordinates (224 women, 226 men). For 95 supervisors, we obtained data from two subordinates, and for 57 supervisors, we obtained data from three or more subordinates. Subordinates were on average 41.36 ( $SD = 10.55$ ) years old, worked 36.74 ( $SD = 10.04$ ) hours per week, and had been working together with their respective supervisor for 3.59 ( $SD = 3.17$ ) years.

### Interviewers

Interviewers were 54 (41 women, 13 men) advanced psychology students who had been trained as interviewers in a one-day frame-of-reference training (see Bernardin & Buckley, 1981; Roch et al., 2012). During this training, interviewers learned about the definitions of task-, relations-, and change-oriented leadership behaviors, and how to evaluate them. Interviewers were on average 29.94 ( $SD = 7.52$ ) years old, had studied psychology as a major for 6.93 ( $SD = 3.07$ ) semesters, and about 62% of interviewers held a Bachelors' degree. About 21% of interviewers had more than ten years of work experience, 59% had between one and ten years of work experience, and 20% had less than one year of work experience.

### Role-players

Role-players were 52 (43 women, 9 men) advanced psychology students who rated supervisors' situational leader effectiveness in two behavioral leadership tasks as part of the leadership assessment program. Role-players had previously participated in a role-player training, in which they received a script for their role and exercised the role-playing several times. Role-players were on average 29.45 ( $SD = 8.06$ ) years old, had studied psychology as a major for 6.76 ( $SD = 2.60$ ) semesters, and about 65% of them held a Bachelors' degree. About 16% of role-players had more than ten years of work experience, 61% had between one and ten years of work experience, and 23% had less than one year of work experience.

### Procedure

Before participating, supervisors provided informed consent that their data could be used for research purposes. At the beginning of the assessment program, supervisors completed an online questionnaire assessing their own leadership behaviors, and their subordinates filled in an online questionnaire which included questions pertaining to their supervisor's leadership behaviors, their own well-being and their affective organizational commitment. Subsequently, supervisors participated in an on-site assessment in which they (a) participated in a structured face-to-face interview assessing different leadership behavior constructs, (b) completed two behavioral leadership tasks, in which role-players rated their situational leader effectiveness, and (c) filled in paper-and-pencil questionnaires including items on their core-self evaluations and emotional intelligence. All assessment instruments were presented in randomized order.

In the structured interview, supervisors responded to seven past-behavior and seven future-behavior questions referring to specific leadership situations. Each interview was administered by a panel of two interviewers and took about 35 min. Interviewers were equipped with an interview guide that provided detailed instructions on how to conduct the interview and that contained all interview questions. After each interview question, interviewers took notes, and then individually rated supervisors' responses. After all interviews were completed, the two interviewers compared and discussed their individual ratings. Interviewers did not have access to supervisors' self-ratings, subordinate ratings, nor to role-players' ratings.

In the two behavioral leadership tasks, supervisors had to take a leadership role and present on a leadership-related topic (i.e., introduce

a new project in a team meeting, or inform subordinates about a corporate reorganization) while two role-players, who took the role of subordinates, asked critical standardized questions. Role-players were instructed to rate how effectively supervisors performed in these tasks. In each behavioral leadership task, supervisors were rated by a different pair of role-players. Thus, in total, each supervisor was rated by four role-players to allow for a reliable assessment of situational leader effectiveness. Supervisor's behavior in the two behavioral leadership tasks was videotaped.

At the end of the on-site assessment, supervisors provided demographic and additional information (e.g., years of leadership experience, and their annual income including salary and any other forms of monetary compensation). Afterwards, they were debriefed regarding the leadership behavior constructs that had been assessed during the assessment program, and they received extensive feedback and suggestions to further improve their leadership skills.

After all other data had been collected, we obtained behavioral codings of leadership behaviors. Therefore, an independent sample of coders watched video recordings of the two behavioral leadership tasks and coded the amount of time that each supervisor engaged in each type of leadership behaviors.

### Measures

#### Supervisor self-ratings

Supervisors rated their own task-oriented leadership behaviors on a 12-item scale, relations-oriented behaviors on a 15-item scale, and change-oriented behaviors on a 12-item scale, all from the Managerial Practices Survey (MPS G16-3; Yukl, 2012; Yukl et al., 2002). Example items for leadership behaviors are "As a supervisor, I clearly explain task assignments and member responsibilities" (task-orientation), "As a supervisor, I show concern for the needs and feelings of individual team members" (relations-orientation), and "As a supervisor, I describe a proposed change or new initiative with enthusiasm and optimism" (change-orientation). Items were rated on a scale from 1 (*not at all*) to 5 (*to a very great extent*).

Core-self evaluations were measured with a 12-item scale from Judge, Bono, Erez, and Locke (2005), and emotional intelligence was measured with 32 items from a scale from Schutte et al. (1998). Example items are "I am confident I get the success I deserve in life" (core-self evaluations) and "I easily recognize my emotions as I experience them" (emotional intelligence). Items were rated on a scale from 1 (*strongly disagree*) to 5 (*strongly agree*). We excluded one item ("I expect that I will do well on most things I try") from the original scale for emotional intelligence because it substantially reduced reliability. Cronbach's alphas for all scales used in this study are provided in Table 1.

#### Subordinate ratings

Subordinates rated their supervisor's task-, relations-, and change-oriented behaviors on the same leadership scales as supervisors rated their own behaviors (MPS G16-3; Yukl, 2012; Yukl et al., 2002). Again, all items were rated on a scale from 1 (*not at all*) to 5 (*to a very great extent*). Given that we posited hypotheses at the supervisor level, we aggregated subordinate ratings for each supervisor and assessed indices of within-group agreement ( $r_{wg(j)}$ ; James, Demaree, & Wolf, 1984, 1993) and consistency across raters based on the one-way random effects model (ICC[1] and ICC[2]; Bartko, 1976; Bliese, 2000). We assumed slightly skewed null distributions to calculate  $r_{wg(j)}$  relying on findings from previous studies that used the same leadership scales (Yukl, Mahsud, Hassan, & Prussia, 2013; Yukl, O'Donnell, & Taber, 2009). Within group agreement ranged from  $r_{wg(j)} = 0.80$  for change-orientation to  $r_{wg(j)} = 0.86$  for relations-orientation leadership, and rater consistency ranged from ICC(1) = 0.19 and ICC(2) = 0.42 for task-orientation to ICC(1) = 0.29 and ICC(2) = 0.55 for relations-orientation. Hence, agreement and consistency were comparable to the

**Table 1**  
Means, standard deviations, and intercorrelations of study variables.

Variables	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8	9	10	11	12
Interview ratings														
1 Overall leadership rating	3.07	0.43	0.91											
2 Task-orientation	3.30	0.50	0.78**	0.85										
3 Relations-orientation	3.38	0.51	0.79**	0.42**	0.86									
4 Change-orientation	2.52	0.57	0.85**	0.50**	0.51**	0.83								
Supervisor demographics														
5 Age	44.95	7.41	0.06	0.09	0.06	0.00	–							
6 Gender <sup>a</sup>	0.59	0.49	0.03	0.03	–0.02	0.05	–0.02	–						
7 Leadership experience	9.32	6.54	0.05	0.03	0.05	0.03	0.59**	0.16	–					
Supervisor self-ratings														
8 Overall leadership rating	3.80	0.41	0.11	0.06	0.10	0.10	–0.02	–0.10	–0.06	0.91				
9 Task-orientation	3.63	0.50	0.02	0.02	0.01	0.02	–0.11	–0.12	–0.22**	0.75**	0.82			
10 Relations-orientation	3.97	0.42	0.15	0.11	0.14	0.10	0.09	–0.20*	0.04	0.79**	0.38**	0.83		
11 Change-orientation	3.80	0.60	0.10	0.02	0.09	0.12	–0.02	0.04	0.04	0.86**	0.42**	0.59**	0.87	
12 Core self-evaluations <sup>b</sup>	3.90	0.39	0.10	0.06	0.03	0.14	–0.01	0.13	0.13	0.08	0.00	0.00	0.17*	0.75
13 Emotional intelligence	3.77	0.32	0.12	0.01	0.07	0.20*	–0.09	–0.13	–0.10	0.52**	0.30**	0.42**	0.52**	0.19*
Subordinate ratings														
14 Overall leadership rating	3.78	0.41	0.13	0.07	0.09	0.15	0.05	–0.04	0.10	0.21*	0.13	0.18*	0.19*	0.01
15 Task-orientation	3.76	0.44	0.11	0.09	0.07	0.09	–0.01	–0.11	0.03	0.16*	0.28**	0.04	0.07	0.04
16 Relations-orientation	3.86	0.47	0.14	0.08	0.11	0.14	0.12	–0.08	0.14	0.20*	0.07	0.29**	0.14	–0.05
17 Change-orientation	3.72	0.52	0.09	0.01	0.05	0.14	0.01	0.06	0.09	0.18*	0.01	0.13	0.26**	0.03
Behavioral codings														
18 Overall leadership rating	151.43	39.50	0.26**	0.24**	0.21**	0.19*	0.24**	0.02	0.07	0.07	–0.02	0.08	0.11	0.14
19 Task-orientation	256.08	78.09	0.11	0.15	0.10	0.03	0.28**	0.01	0.12	0.04	–0.01	0.10	0.03	0.06
20 Relations-orientation	97.30	39.47	0.33**	0.23**	0.26**	0.30**	0.18*	–0.05	0.04	0.09	0.04	0.06	0.12	0.17*
21 Change-orientation	100.90	42.84	0.22**	0.18*	0.16	0.19*	–0.04	0.08	–0.06	0.02	–0.08	–0.03	0.13	0.12
Criteria														
22 Annual income <sup>c</sup>	11.93	0.38	0.23*	0.17	0.23*	0.17	0.25**	0.23*	0.32**	–0.13	–0.32**	–0.02	0.00	0.14
23 Leader effectiveness	5.20	0.84	0.23**	0.30**	0.11	0.16	–0.02	–0.07	–0.06	0.03	–0.05	0.07	0.05	0.16*
24 Subordinates' well-being	4.53	0.50	0.25**	0.21**	0.16*	0.24**	0.12	0.00	0.13	0.08	0.08	0.00	0.09	0.17*
25 Subordinates' commitment	5.08	0.74	0.21**	0.16	0.04	0.30**	0.13	0.11	0.17*	0.09	0.06	0.00	0.13	0.07

Variables	13	14	15	16	17	18	19	20	21	22	23	24	25
Supervisor self-ratings													
13 Emotional intelligence	0.83												
Subordinate ratings													
14 Overall leadership rating	0.09	0.96											
15 Task-orientation	0.02	0.80**	0.91										
16 Relations-orientation	0.07	0.90**	0.58**	0.93									
17 Change-orientation	0.14	0.90**	0.54**	0.77**	0.92								
Behavioral codings													
18 Overall leadership rating	0.01	–0.04	–0.05	–0.02	–0.04	–							
19 Task-orientation	–0.04	–0.05	–0.04	0.00	–0.08	0.83**	–						
20 Relations-orientation	0.15	0.03	0.01	0.03	0.04	0.58**	0.16	–					
21 Change-orientation	–0.04	–0.06	–0.07	–0.10	0.01	0.71**	0.34**	0.39**	–				
Criteria													
22 Annual income <sup>c</sup>	–0.05	0.02	–0.03	0.03	0.06	0.37**	0.28**	0.18	0.32**	–			
23 Leader effectiveness	0.14	0.09	0.07	0.09	0.06	0.37**	0.22**	0.31**	0.35**	0.22*	0.95		
24 Subordinates' well-being	0.16*	0.37**	0.31**	0.29**	0.36**	0.04	0.03	0.04	0.02	0.11	0.07	0.78	
25 Subordinates' commitment	–0.02	0.41**	0.32**	0.33**	0.40**	–0.01	0.01	0.01	–0.06	0.10	0.11	0.41**	0.83

Note. *N* = 152; Overall leadership rating = mean of task-, relations-, and change-orientation; Cronbach's alphas are presented in the table's diagonal.  
*n* = 114.

<sup>a</sup> 0 = female, 1 = male.

<sup>b</sup> *n* = 151.

<sup>c</sup> Variable has been log-transformed;

\* *p* < .05.

\*\* *p* < .01.

values reported in the literature (Woehr, Loignon, Schmidt, Loughry, & Ohland, 2015).

Subordinates rated their general well-being on a 5-item scale from the World Health Organization (WHO-5; Topp et al., 2015), and their affective organizational commitment on the revised 6-item scale from Meyer and Allen (1997). Example items are “During the last two weeks, I have felt active and vigorous” (general well-being), and “This organization has a great deal of personal meaning for me” (affective

commitment). Items for general well-being were rated on a scale from 1 (*never*) to 6 (*all the time*), and items for affective commitment were rated on a scale from 1 (*strongly disagree*) to 7 (*strongly agree*). Again, we examined agreement within and consistency across subordinates before averaging their ratings. Based on previous research using these variables, we assumed slightly skewed null distributions to calculate  $r_{wg(j)}$  (Bernecker, Herrmann, Brandstätter, & Job, 2017; Meyer, Allen, & Smith, 1993). Regarding general well-being, within group agreement

was  $r_{wg(j)} = 0.71$  and, thus, comparable to the mean values for within-group agreement as reported in the literature (Woehr et al., 2015). As expected, rater consistency was rather low with  $ICC(1) = 0.06$  and  $ICC(2) = 0.16$ , given that subordinates evaluated their general well-being which can be influenced by many factors (e.g., Nielsen et al., 2017). For affective organizational commitment, within group agreement ( $r_{wg(j)} = 0.67$ ), and rater consistency ( $ICC[1] = 0.19$  and  $ICC[2] = 0.41$ ) were comparable to the levels of agreement and consistency reported in the literature (Woehr et al., 2015).

#### Interview ratings

Interview development proceeded in six steps. First, interview items were developed based on the critical incident technique (Flanagan, 1954). We used 14 out of 32 critical incidents that had been collected in a study by Peus et al. (2013) via focus group discussions with supervisors. We selected those critical incidents from Peus et al. (2013) that referred to (a) situations in which supervisors interacted with their subordinates, (b) challenging situations in which more than one type of leadership behavior can lead to a desired outcome, and (c) situations that are likely to be a common experience for most supervisors.

Second, these 14 identified situations were enriched so that each situation contained cues that can activate task-, relations-, and change-oriented leadership behaviors. Thus, each situation was designed to capture three leadership behavior constructs simultaneously. This is because situations that are indicative of effective leadership are complex and thus offer a wide array of behavioral reactions that are associated with different kinds of leadership behaviors (e.g., Mumford, Zaccaro, Harding, Jacobs, & Fleishman, 2000).

Third, out of the 14 identified situations, half of the situations were used to write past-oriented interview items (i.e., they were framed to ask for supervisors' past experiences), and the other half were used to write future-oriented interview items (i.e., they were framed into hypothetical situations). Past-oriented and future-oriented questions were comparable in length and in complexity, and they were administered by the same interviewers and in the same manner.

Fourth, we developed rating scales with behavioral examples for task-, relations-, and change-oriented leadership behaviors for each interview item. Thus, each interview item together with its three rating scales formed one interview question. Behavioral examples for the rating scales were adapted items from established leadership questionnaires, specifically the Leadership Behavior Description Questionnaire (LBDQ; R. M. Stogdill, Goode, & Day, 1962), the Transformational Leadership Inventory (TLI; Podsakoff et al., 1990), and the Managerial Practices Survey (MPS G16-3; Yukl, 2012; Yukl et al., 2002). For each interview item, interviewers evaluated each leadership behavior construct on five-point rating scales. Given that interviewers provided three ratings (i.e., one rating of task-, relations-, and change-oriented behavior respectively) for each of the 14 interview items (i.e., seven past-behavior and seven future-behavior items), the interview comprised 42 rating scales in total.

Fifth, we thoroughly revised the rating scales by adapting them more closely to the specific contexts of the respective interview item. Examples of past-oriented and future-oriented interview questions with their respective rating scales can be found in the appendix.

Sixth, we asked seven subject matter experts (SMEs) to provide ratings on the content validity of the developed interview. SMEs were researchers in the field of I/O psychology with a focus on leadership and/or assessment methods. SMEs were provided with the definitions of the leadership behavior constructs and received the full leadership interview, but the labels of the ratings scales (i.e., the names of the leadership constructs) were blinded. First, SMEs were asked to identify the respective leadership construct from Yukl's taxonomy for each of the rating scales. Results show that all SMEs accurately identified the intended leadership construct for each single rating scale. Second, SMEs rated the extent to which each of the rating scales (whose labels were still blinded) were suited to measure each of the three leadership

constructs; scales ranged from 1 (*strongly disagree*) to 5 (*strongly agree*). On average, SMEs indicated that the rating scales were well suited to assess the leadership construct that the respective rating scale intended to measure ( $M = 4.86$ ,  $SD = 0.17$ ), while rating scales were less suited to assess the two leadership constructs that the respective rating scale did not intend to measure ( $M = 1.33$ ,  $SD = 0.24$ ). Third, SMEs were asked to rate the overall interview, specifically whether interview questions represented relevant leadership situations and were suited to assess the intended leadership constructs. Again, SMEs made ratings on a scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). Their ratings illustrated that the interview questions generally represented relevant leadership situations ( $M = 4.71$ ,  $SD = 0.45$ ), and that the interview questions were well suited to assess the three leadership constructs ( $M = 4.86$ ,  $SD = 0.35$ ).

Finally, each interview was administered by two interviewers. Thus, for one supervisor, the same panel of two interviewers rated all 14 interview questions. We examined agreement within and consistency across interviewers before averaging their ratings. Both within group agreement ( $r_{wg(j)} = 0.98$ ), and rater consistency ( $ICC[1] = 0.94$  and  $ICC[2] = 0.97$ ) were high (Woehr et al., 2015). In addition, the mean correlation between interviewers ( $r = 0.66$ ) was comparable to the interviewer reliability reported for previous leadership interviews (Huffcutt, Weekley, et al., 2001). We averaged interviewers' combined ratings across interview questions to obtain one score for each leadership behavior construct and one score for the overall interview across the three leadership behavior constructs respectively.

#### Role-player ratings

Role-players rated supervisors' situational leader effectiveness with a 4-item scale from Van Knippenberg and Van Knippenberg (2005) in two behavioral leadership tasks. An example item is "I think that this supervisor works effectively". Items were rated on a scale from 1 (*strongly disagree*) to 7 (*strongly agree*). To ensure the reliability of these ratings, multiple role-players (i.e., two role-players in each of the two leadership tasks) rated each supervisor. Both within group agreement ( $r_{wg(j)} = 0.85$ ), and rater consistency ( $ICC[1] = 0.49$  and  $ICC[2] = 0.79$ ) corresponded to values reported in the literature (Woehr et al., 2015). Thus, we averaged ratings across role-players and items to form one score of situational leader effectiveness.

#### Behavioral codings

An independent sample of coders recorded supervisor's task-, relations-, and change-oriented behaviors based on videos of the two simulated leadership tasks. Coders used the INTERACT video coding software (Mangold International, 2010), similar to other behavioral research (e.g., Kauffeld & Meyers, 2009; Lehmann-Willenbrock & Allen, 2014). Using this software, coders watched the videos and each time they identified a relevant behavior, they pressed a computer key that was programmed to represent the respective leadership category. At the same time, coders marked the time period (i.e., how long supervisors engaged in each behavior). Hence, behavioral codings obtained in this study represent the amount of time that each supervisor engaged in each leadership behavior.

Coders were three graduate students in psychology who were blind to all other data collected in this study. Before coding the videos, they had to complete 10 h of training. To assess the reliability of behavioral codings, the first 30 videos of the sample were independently coded by all three coders. We calculated ICCs for absolute agreement using the two-way random effects model (McGraw & Wong, 1996). Coder consistency ranged from  $ICC(1) = 0.69$  and  $ICC(2) = 0.87$  for change-orientation to  $ICC(1) = 0.92$  and  $ICC(2) = 0.97$  for task-orientation. Given that coder consistency was good and in line with previous research, coders then proceeded to code different videos so that each video was coded by one coder (Kauffeld & Lehmann-Willenbrock, 2012; McFarland, Yun, Harold, Viera, & Moore, 2005).



## Results

**Hypothesis 1** stated that a factor model specifying task-, relations-, and change-oriented leadership as distinct factors would best represent the internal data structure of interview ratings. To test this hypothesis, we conducted a set of CFAs and tested four potential models: Model I was a common multitrait-multimethod model with three correlated dimension factors (i.e., the leadership behavior constructs) and two correlated method factors (i.e., past-behavior questions and future-behavior questions). We chose to start with this model because it most adequately reflects how the interview was designed. Model II included only the three correlated dimensions factors and was nested in Model I. Model III included only the two correlated method factors and was also nested in Model I. Model IV included the two correlated method factors and one overall dimension factor. Thus, Model IV was not nested in Model I. We chose to also test this model given that interview ratings are often averaged across different (leadership) dimensions to form one overall score (e.g., [Krajewski et al., 2006](#)).

Before testing these models, we parceled interview ratings in order to reduce model complexity and to achieve an acceptable ratio of sample size to estimated parameters ([Bentler & Chou, 1987](#)). The interview consisted of 42 ratings (i.e., interview ratings of three leadership behavior constructs each rated in seven past-behavior and seven future-behavior questions). Thus, we had seven ratings of each leadership construct for each interview type. We assigned these seven ratings per leadership construct and interview type to three parcels.<sup>1</sup> Therefore, the analyses of all factor models were based on 18 parcels in total: Each dimension factor (i.e., latent leadership behavior construct) was measured with six parcels and each method factor (i.e., latent type of interview question) was measured with nine parcels.

We computed factor models with Mplus version 7.11 ([Muthén & Muthén, 2013](#)). Model I fit the data well,  $\chi^2(113) = 136.65$ ,  $p = .064$ , TLI = 0.97, CFI = 0.98, RMSEA = 0.04, SRMR = 0.04, according to standards for model evaluation (e.g., [Antonakis, Bendahan, Jacquart, & Lalive, 2010](#); [Schermelleh-Engel, Moosbrugger, & Müller, 2003](#)). Model comparisons using the chi-squared test further revealed that Model I fit the data significantly better than Model II,  $\Delta\chi^2(19) = 72.29$ ,  $p < .001$ , and that Model I fit the data significantly better than Model III,  $\Delta\chi^2(21) = 296.59$ ,  $p < .001$ . We could not test directly whether Model I fit the data better than Model IV given that Model IV was not nested in Model I. However, Model IV did not show good fit,  $\chi^2(116) = 250.11$ ,  $p < .001$ , and the AIC was lower for Model I (5068.37) than for Model IV (5175.83). Therefore, the model that provided the best fit to the data was Model I, which incorporated the three leadership behavior constructs as correlated dimension factors, and the two types of interview questions as correlated method factors. Consequently, **Hypothesis 1** was supported. [Fig. 1](#) shows the best fitting model.

Similar to previous interview studies (e.g., [Ingold, Kleinmann, König, & Melchers, 2015](#); [Morgeson, Reider, & Campion, 2005](#)), we averaged ratings across question types for each leadership behavior construct before testing further hypotheses and research questions. We decided on this approach given that ratings of past-oriented and future-oriented interview questions showed high correlations. Specifically, observed correlations between question types were  $r = 0.62$  (for task-orientation),  $r = 0.62$  (for relations-orientation), and  $r = 0.63$  (for change-orientation) which exceeded findings from previous interview studies (i.e., the meta-analytic correlation between construct-matched past-oriented and future-oriented interview questions is  $r = 0.40$ ; [Culbertson et al., 2017](#)).

<sup>1</sup> We chose a rational (i.e., non-empirical) technique for assigning interview ratings to parcels ([Landis, Beal, & Tesluk, 2000](#)). Specifically, we assigned the 42 indicators (i.e., interview ratings) to 18 parcels based on the time sequence of interview questions. To test the stability of results, we additionally computed the hypothesized model by using a single-factor (i.e., empirical) parceling technique ([Landis et al., 2000](#)), which yielded results comparable to those from the rational parceling technique.

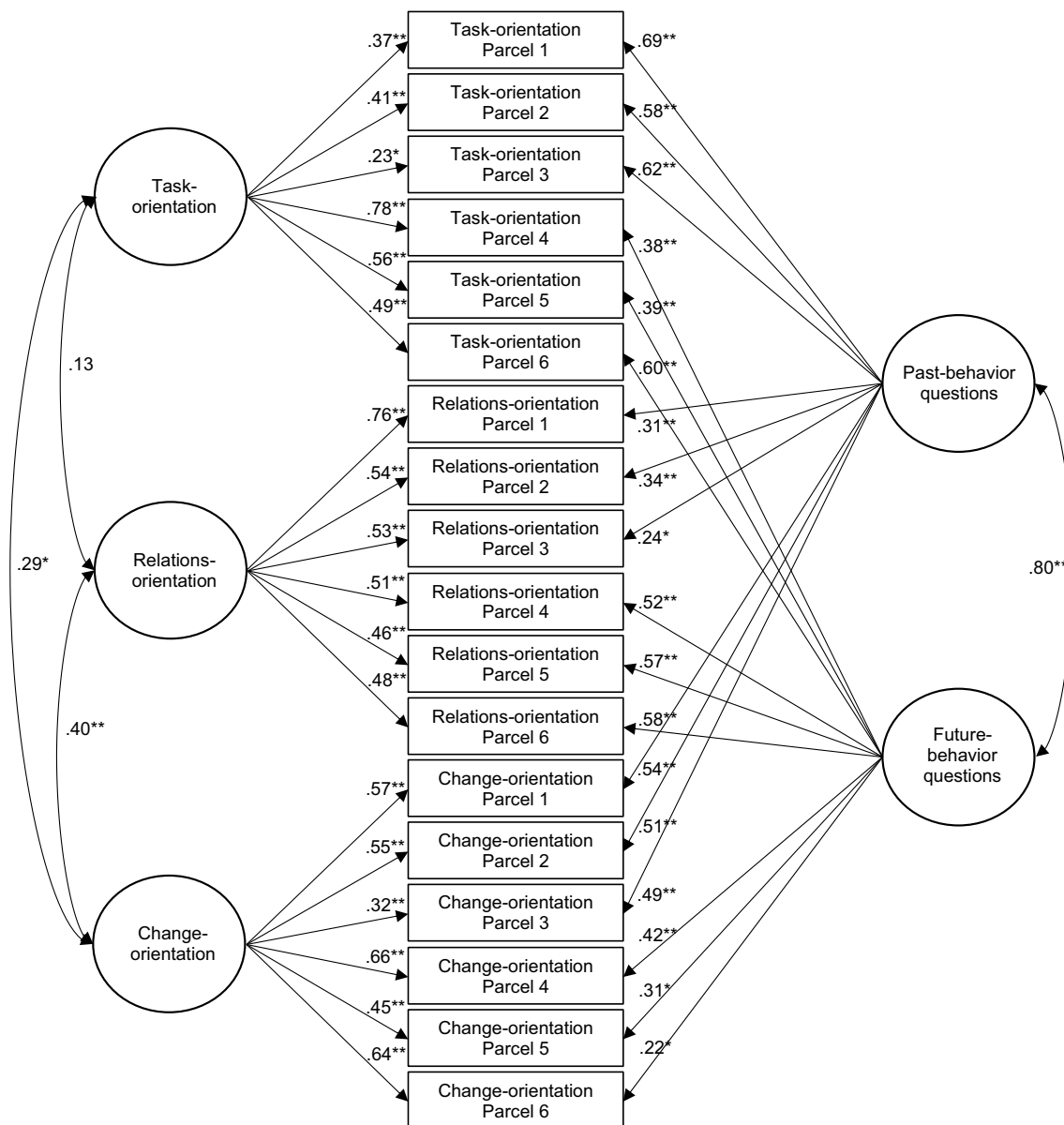
**Hypothesis 2** posited that supervisors who achieved a higher overall interview score (i.e., interview ratings averaged across the three leadership behavior constructs) would have a higher reported annual income. Therefore, the overall interview score should have incremental validity over and above characteristics of the organization, leader traits, and questionnaire-based self-ratings and subordinate ratings of leadership behaviors. Given that skewness and kurtosis scores (4.02 and 19.91) of annual income indicated a positively skewed distribution, we log-transformed the income variable (e.g., [Tabachnick & Fidell, 2007](#)). The overall interview score correlated positively with supervisors' log-transformed reported annual income ( $r = 0.23$ ,  $p = .004$ ), as shown in [Table 1](#). The overall interview score was also correlated positively with annual income when the variable was not log-transformed ( $r = 0.20$ ,  $p = .038$ ). As shown in [Table 2](#) (see results for Model 2), hierarchical regression analysis further indicated that the overall interview score explained a significant proportion of variance in supervisors' log-transformed reported annual income beyond characteristics of the organization, leader traits, and questionnaire-based self-ratings and subordinate ratings of leadership behaviors,  $\Delta R^2 = 0.04$ ,  $F(1, 104) = 6.26$ ,  $p = .014$ . Hence, **Hypothesis 2** was supported. To further determine the relative contribution of all predictors in explaining variance in annual income, we conducted relative weights analyses with the *relaimpo* package for the R environment ([Grömping, 2006](#)). This type of analysis is recommended when predictors are intercorrelated ([Johnson, 2000](#)), which is often the case when analyzing leadership variables. As can also be seen in [Table 2](#), relative weights analysis revealed that the overall interview score accounted for 20.7% of the variance explained in annual income, while the other predictors explained between 1.0% (emotional intelligence) and 24.7% (leadership experience) of variance.

**Hypothesis 3** assumed that supervisors achieved a higher overall interview score would obtain higher ratings of situational leader effectiveness in behavioral leadership tasks. As can be seen in [Table 1](#), the overall interview score correlated significantly with ratings of situational leader effectiveness ( $r = 0.23$ ,  $p = .004$ ). Hierarchical regression analysis further indicated that the overall interview score explained a significant proportion of variance in ratings of situational leader effectiveness beyond leader traits (including demographic variables), questionnaire-based self-ratings and subordinate ratings of leadership behaviors,  $\Delta R^2 = 0.04$ ,  $F(1, 143) = 7.38$ ,  $p = .007$ , see [Table 2](#) (results for Model 4). As such, **Hypothesis 3** was supported. Furthermore, relative weights analysis demonstrated that 45.4% of the variance explained in ratings of situational leader effectiveness could be attributed to the interview overall score, while the other predictors explained between 0.8% (age) and 20.6% (core self-evaluation) of the variance.

**Hypothesis 4** stated that supervisors with a higher overall interview score would have subordinates reporting a higher level of general well-being. In support of this, the interview score correlated significantly with subordinates' general well-being ( $r = 0.25$ ,  $p = .002$ ). Further analyses demonstrated that the overall interview score explained a significant proportion of variance in subordinates' well-being beyond leader traits (including demographic variables), questionnaire-based self-ratings and subordinate ratings of leadership behaviors,  $\Delta R^2 = 0.03$ ,  $F(1, 143) = 5.806$ ,  $p = .017$ , as shown in [Table 2](#) (see results for Model 6). Accordingly, **Hypothesis 4** was supported. In addition, relative weights analysis demonstrated that 20.1% of the variance explained in subordinates' well-being could be attributed to the overall interview score, while the other predictors explained between 0.1% (gender) and 53.8% (overall score of subordinate ratings of leadership behaviors) of the variance.<sup>2</sup>

**Hypothesis 5** expected supervisors with a higher overall interview score to have subordinates with higher levels of affective organizational

<sup>2</sup> Please note that one reason why subordinate ratings of leadership behaviors explained a large amount of variance in general well-being and affective organizational commitment is that these outcomes were also rated by subordinates which implies some common-method bias ([Podsakoff, MacKenzie, Lee, & Podsakoff, 2003](#)).



**Fig. 1.** Confirmatory multitrait-multimethod factor analysis of interview ratings. Factor loadings on trait factors (i.e., leadership behaviors) were 0.53 on average and factor loadings on method factors (i.e., question types) were 0.47 on average.  $N = 152$ . \* $p < .05$ . \*\* $p < .01$ .

commitment. In line with this, the interview score correlated significantly with subordinates' affective commitment ( $r = 0.21$ ,  $p = .009$ ). As shown in Table 2 (see results for Model 8), analyses demonstrated that the overall interview score explained a significant proportion of variance in subordinates' affective commitment beyond leader traits (including demographic variables), questionnaire-based self-ratings and subordinate ratings of leadership behaviors,  $\Delta R^2 = 0.02$ ,  $F(1, 143) = 4.233$ ,  $p = .040$ . Hence, Hypothesis 5 was supported. Relative weights analysis further revealed that 15% of the variance explained in subordinates' affective commitment could be attributed to the overall interview score, while the other predictors explained between 1.2% (core self-evaluations) and 66.7% (overall score of subordinate ratings of leadership behaviors).<sup>2</sup>

Following best practice recommendations on handling control

variables in leadership research (e.g., Bernerth, Cole, Taylor, & Walker, 2018), we additionally tested Hypotheses 2 to 5 (a) without demographic variables (age, gender, leadership experience) and (b) without demographic variables and self-rated leader traits (core-self-evaluations, emotional intelligence). In each case, the interview overall score remained a significant predictor of all examined leadership outcomes.

Finally, as a research question, we explored whether interview ratings of leadership behavior constructs relate to other measures of leadership behavior (i.e. self-ratings, subordinate ratings, and behavioral codings of the same constructs, see Table 1). Interview ratings showed small, non-significant correlations with questionnaire-based self-ratings for task-orientation,  $r = 0.02$ ,  $p = .818$ , relations-orientation,  $r = 0.14$ ,  $p = .079$ , and change-orientation,  $r = 0.12$ ,  $p = .157$ . These results correspond to previous research which found little or no

**Table 2**

Hierarchical regression and relative weights analyses predicting supervisors' annual income, leader effectiveness, and subordinates' general well-being and affective organizational commitment.

Variables	Model 1				Model 2			
	$\beta$	SE	RW	%RW	$\beta$	SE	RW	%RW
Supervisors' annual income ( $N = 114$ )								
Size of the organisation <sup>b</sup>	0.13	0.09	0.011	6.0	0.13	0.09	0.012	4.9
Industry <sup>c</sup>	−0.07	0.11	0.010	5.2	−0.03	0.10	0.008	3.3
Leader age	0.12	0.13	0.035	18.1	0.10	0.13	0.034	14.2
Leader gender <sup>a</sup>	0.18	0.09	0.037	19.3	0.17	0.09	0.035	14.8
Leadership experience	0.19	0.13	0.057	30.0	0.21	0.13	0.059	24.7
Leader core self-evaluations	0.12	0.10	0.016	8.4	0.11	0.10	0.014	6.1
Leader emotional intelligence	0.03	0.10	0.002	1.0	0.01	0.10	0.002	1.0
Overall score self-rated leadership	−0.18	0.11	0.019	10.0	−0.20	0.11	0.022	9.2
Overall score subordinate-rated leadership	0.09	0.09	0.004	2.0	0.07	0.09	0.003	1.2
Overall score interview ratings of leadership					0.21*	0.08	0.049	20.7
$R^2$	0.19**				0.24**			
$\Delta R^2$					0.05*			
Ratings of leader effectiveness ( $N = 151$ )								
Variables	Model 3				Model 4			
	$\beta$	SE	RW	$\beta$	$\beta$	SE	RW	$\beta$
Leader age	0.04	0.10	0.001	1.3	0.02	0.10	0.001	0.8
Leader gender <sup>a</sup>	−0.08	0.08	0.008	11.5	−0.09	0.08	0.008	7.4
Leadership experience	−0.11	0.10	0.006	9.7	−0.10	0.10	0.007	5.9
Leader core self-evaluations	0.16	0.08	0.026	38.8	0.15	0.08	0.023	20.6
Leader emotional intelligence	0.14	0.09	0.020	29.8	0.12	0.09	0.017	15.4
Overall score self-rated leadership	−0.07	0.09	0.002	3.2	−0.08	0.09	0.002	2.1
Overall score subordinate-rated leadership	0.07	0.08	0.004	5.8	0.04	0.08	0.003	2.3
Overall score interview ratings of leadership					0.21**	0.08	0.061	45.4
$R^2$	0.07				0.11*			
$\Delta R^2$					0.05**			
Subordinates' general well-being ( $N = 151$ )								
Variables	Model 5				Model 6			
	$\beta$	SE	RW	%RW	$\beta$	SE	RW	%RW
Leader age	0.10	0.09	0.009	4.7	0.08	0.09	0.008	3.6
Leader gender <sup>a</sup>	0.02	0.08	0.000	0.1	0.01	0.08	0.000	0.1
Leadership experience	0.03	0.10	0.009	4.4	0.03	0.10	0.008	3.6
Leader core self-evaluations	0.13	0.08	0.021	10.8	0.12	0.08	0.019	8.4
Leader emotional intelligence	0.17	0.09	0.022	11.3	0.15	0.09	0.020	8.6
Overall score self-rated leadership	−0.10	0.09	0.004	2.1	−0.10	0.09	0.004	1.8
Overall score subordinate-rated leadership	0.37***	0.08	0.130	66.7	0.35***	0.08	0.122	53.8
Overall score interview ratings of leadership					0.18*	0.08	0.046	20.1
$R^2$	0.20***				0.23***			
$\Delta R^2$					0.03*			
Subordinates' affective commitment ( $N = 151$ )								
Variables	Model 7				Model 8			
	$\beta$	SE	RW	%RW	$\beta$	SE	RW	%RW
Leader age	0.07	0.09	0.008	4.0	0.06	0.09	0.007	3.3
Leader gender <sup>a</sup>	0.10	0.08	0.010	4.8	0.09	0.08	0.009	4.1
Leadership experience	0.06	0.10	0.013	6.8	0.06	0.10	0.013	5.9
Leader core self-evaluations	0.05	0.08	0.003	1.6	0.04	0.08	0.003	1.2
Leader emotional intelligence	−0.06	0.09	0.002	1.0	−0.08	0.09	0.003	1.3
Overall score self-rated leadership	0.05	0.09	0.006	3.1	0.05	0.09	0.006	2.6
Overall score subordinate-rated leadership	0.39***	0.08	0.154	78.5	0.38***	0.08	0.146	66.7
Overall score interview ratings of leadership					0.16*	0.08	0.033	15.0
$R^2$	0.20***				0.22***			
$\Delta R^2$					0.02*			

Note. Overall score = mean of task-, relations-, and change-orientation; RW = relative weights of predictors summing up to  $R^2$ ; %RW = percentages of relative weights.

<sup>a</sup> 0 = female, 1 = male.

<sup>b</sup> 0 = less than 250 employees, 1 = 250 or more employees.

<sup>c</sup> 0 = private sector, 1 = public sector.

\*  $p < .05$ .

\*\*  $p < .01$ .

\*\*\*  $p < .001$ .

convergence between interview ratings and self-ratings (Atkins & Wood, 2002; Van Iddekinge et al., 2004; Van Iddekinge et al., 2005). Correlations between interview ratings and subordinate ratings were also non-significant for task-orientation,  $r = 0.09$ ,  $p = .275$ , relations-orientation,  $r = 0.11$ ,  $p = .173$ , and change-orientation,  $r = 0.14$ ,  $p = .076$ . Still, the present leadership interview showed stronger convergence with subordinate ratings than the only previous study examining these two types of leadership measures (Atkins & Wood, 2002). Correlations between interview ratings and behavioral codings were higher and significant for relations-orientation,  $r = 0.26$ ,  $p = .001$ , and change-orientation,  $r = 0.19$ ,  $p = .022$ , but non-significant for task-orientation,  $r = 0.15$ ,  $p = .066$ . Thus, interview ratings descriptively showed the highest convergence with behavioral codings of leadership constructs.

## Discussion

Leadership research has called for new approaches to measuring leadership behaviors (Antonakis et al., 2016; DeRue et al., 2011; Hunter et al., 2007). The present study addressed this call by thoroughly investigating the potential of the structured interview method for assessing leadership behavior constructs, thereby providing initial validity evidence.

As a first major contribution, the present study showed that the overall score of an interview designed to assess leadership behavior constructs predicts a variety of outcomes including supervisors' annual income and role-players' ratings of situational leader effectiveness, as well as subordinates' well-being and affective organizational commitment. Thereby, this study expands previous research on the validity of structured interviews for leadership positions which so far has predominantly focused on supervisors' performance as outcome measure (Huffcutt, Weekley, et al., 2001; Krajewski et al., 2006) while neglecting subordinate outcomes. The present finding is of particular practical interest because it implies that a structured leadership interview assessing established leadership constructs can help to identify those leaders who will have a positive impact on their subordinates.

In predicting these different outcomes, interview ratings further explained variance over and above questionnaire-based self-ratings and subordinate ratings of leadership behaviors. This highlights the benefits of a context-specific assessment of leadership behaviors. Both previous research on situational judgment tests (SJTs; Peus et al., 2013) and this study support the assumption that situational assessments (i.e., providing a specific situation as a framing for rating leadership behaviors) explain variance in leadership outcomes over traditional assessments of leadership that do not refer to a specific context (i.e., conventional leadership questionnaires).

As a second major contribution, this study is among the first to find that the data structure of interview ratings reflects the interview dimensions (i.e., constructs) that the interview was designed to measure. The main reason for this finding may be that the present interview was designed to assess broad and well-defined constructs from leadership research (Yukl, 2012). Consequently, this finding provides new opportunities for developing further structured interviews to assess other relevant leadership constructs (e.g., instrumental leadership, ethical leadership, or charismatic leadership; Antonakis et al., 2016; Brown, Treviño, & Harrison, 2005). Broadly speaking, the structured interview can be an appealing option in all contexts where conventional leadership questionnaires cannot be used such as leader selection (due to self-raters' motivational biases and given that ratings from subordinates are often simply not available) or in certain research settings (where common source variance between the assessed leadership behaviors

and leadership outcomes may limit predictive inference; see for example Antonakis et al., 2010).

As a third and final contribution, the present findings improve our understanding of the interview method by shedding light on interrelations between interview ratings and other leadership measures. Specifically, this study is the first to examine how interview ratings of leadership behavior constructs relate to self-ratings, subordinate ratings and behavioral codings of the same leadership constructs. Empirically, we found that interview ratings correspond to behavioral codings, but hardly to self-ratings and subordinate ratings of leadership.

A conceptual explanation for these less intuitive findings at first sight draws from the distinction between maximum and typical performance (Sackett, Zedeck, & Fogli, 1988). Applying the definitions of these performance types to leadership behavior, maximum performance is how supervisors interact with their subordinates when they devote full attention and effort to their leadership role, while typical performance is how supervisors interact with their subordinates on a regular basis (see also Klehe & Anderson, 2007; Sackett, 2007). Following Sackett et al. (1988), individuals are likely to demonstrate maximum performance when they are aware that they are being evaluated and instructed to maximize their effort, and when the evaluation occurs over a short time period so that the individual remains fully focused on their performance. These criteria seem to apply to both interview ratings and behavioral codings of leadership behaviors in simulated situations: In both contexts, supervisors were explicitly evaluated by interviewers/role-players who were physically present in the situation. In addition, both measures referred to leadership behaviors within a specific time frame. On the other hand, self-ratings and subordinate ratings of leadership behaviors seem to tap more into typical performance, given respondents refer to long-term behavioral tendencies and are not instructed to evaluate leadership behavior in a specific situation. Consequently, interview ratings (and also behavioral codings) may be more likely to capture how supervisors act when they are at their best behavior and therefore only modestly correspond to self-ratings and subordinate ratings which describe more typical behaviors.

When it comes to predicting leadership outcomes comprehensively, it may be of interest to assess both maximum and typical leadership behavior. For example, the present findings show that more distal leadership outcomes such as subordinate well-being are predicted by both interview ratings (as maximum performance measure) and subordinate ratings (as typical performance measure). Conceptually, this makes sense when considering that subordinates are affected by how the supervisor behaves in critical situations (when maximum effort is needed) and by how the supervisor behaves in everyday work life situations (when maximum effort is not needed).

## Practical implications

Overall, this study provides encouraging evidence for the use of structured interview method in the field of leader selection and development. Regarding leader selection, the finding that interview ratings demonstrate incremental validity beyond leader traits, self-ratings and subordinate ratings of leadership behaviors suggests that the costs of conducting an interview (e.g., training interviewers, time required for administering interviews as compared to administering questionnaires) may be outweighed by its benefits to validity.

Regarding implications for leader development, the most important finding is that interview ratings are not redundant to but may meaningfully complement other leadership measures. In a developmental program, it is often required to provide supervisors with differentiated feedback on their behaviors. In this context, interview ratings of



leadership behaviors could offer an additional perspective to self-ratings and ratings from other sources (i.e., ratings from subordinates, peers, and superiors as part of a 360° feedback intervention).

### Limitations

This study is not without limitations. While the interview score predicts a wide range of leadership outcomes, we cannot infer from the results how interview ratings relate to supervisors' behavior in their everyday work-life. Results illustrated that those supervisors who obtain higher ratings in the leadership interview had a higher income (with or without considering demographic controls and characteristics of the organization), engaged actively in more leadership behaviors in simulated leadership tasks, were perceived as more effective in these leadership tasks by role-players, and had subordinates who reported higher levels of general well-being and affective organizational commitment. At the same time, interview ratings did not reflect how supervisors perceived themselves or their subordinates' perceptions of supervisors' leadership behaviors. Additionally, the study design did not allow us to record and code how supervisors interact with their actual subordinates at the workplace. Hence, research is warranted to understand how interview ratings relate to how leaders actually behave at their specific jobs.

In addition, we would like to note that measurement error (i.e., using variables that are not perfectly reliable) might have biased the results to some extent (e.g., [Ree & Carretta, 2006](#)). In our analyses, we did not account for error variance when we parceled interview ratings in CFAs and when we predicted leadership outcomes without modeling predictors as latent variables. Thus, some caution is warranted as the present results could slightly overestimate or underestimate the true relationship of interview ratings with leadership outcomes (e.g., [Bollen, 1989](#)). Still, we decided for this more practical analytic approach due to the complexity of the interview data structure and due to a limited sample size (e.g., [Bentler & Chou, 1987](#)).

Finally, self-selection bias could potentially threaten the generalizability of our findings. It is possible that only those supervisors who already invest many resources in being “good leaders” might have participated in the leadership assessment program. However, we primarily recruited supervisors through their organizations (e.g., by contacting the HR departments of local organizations). These organizations often decided to have all their supervisors participate in the program, or the HR department decided (together with the respective supervisors) who would benefit most from the program and should therefore

participate. Consequently, supervisors took part in the present study for various reasons, which lessens concerns about self-selection biases.

### Future research

We see the present study as a promising starting point to intertwine the research avenues of interview and leadership research. In general, more diversity in the assessment of leadership behavior constructs may allow for a more differentiated understanding of how leaders behave and why they do so. The present study has shown that it is a fruitful approach to use structured interviews as one complementary approach to traditional leadership questionnaires. We suggest that future research adapts and explores the potential of further assessment instruments from personnel selection and development (i.e., virtual interviews, assessment center exercises, video-based SJTs, etc.) for measuring constructs from leadership theory that are typically assessed via questionnaires. This would equip research with a broader choice of tools for assessing leaders.

Along these lines, we encourage future research to systematically compare different approaches to measuring leadership behavior to deepen our understanding about how leadership measures function. In particular, comparing measures that vary only with regard to one specific methodological factor such as the information source (e.g., whether the same interview questions are rated by interviewers or by supervisors or by subordinates) may allow for examining which features of the respective measurement approach drive (a) convergence between measures and (b) the validity of different measures. In other words, examining this would generate more detailed knowledge on why certain measures hardly correspond to each other (e.g., interview ratings and self-ratings; see also [Van Iddekinge et al., 2004](#)) and why certain measures are more predictive of leadership outcomes than others.

### Conclusion

Evidence from this study illustrates that structured interviews help to identify leaders who have a positive impact on their subordinates. In addition, interview ratings of leadership behavior constructs predicted criteria beyond a number of other relevant predictors including self-ratings and subordinate ratings of the same leadership constructs. Thus, the interview method shows potential to meaningfully complement existing leadership measures in research and practice.

## Appendix A

## Example 1 of a past-behavior interview question

“Think of a situation in which your employees were skeptical about a new work process you had to introduce and that was necessary for the effectiveness of the organization. Please describe both the situation and your exact response in detail.”

Notes:

Please rate the extent of the respective leadership behavior construct by ticking the appropriate number:

	1	2	3	4	5	
<b>Relations-orientation</b>	Shows no interest in employees' concerns; Does not offer support; Does not allow employees to participate in the design of new work processes		Acknowledges employees' concerns; Offers employees support where necessary; Allows employees to participate in the design of new work processes			Acknowledges employees' concerns individually; Offers employees comprehensive support with the implementation of new work processes; Engages employees in designing new work processes
<b>Task-orientation</b>	Does not explain the new process; States no expectations; Does not check compliance with the new process		Informs employees about the new process; Addresses what is expected of the employees; Checks compliance with the new process if necessary			Develops a concrete plan to implement the new process; Provides clear instructions; Analyzes the problem; Clearly communicates what is expected of the employees; Consistently checks compliance with the new process
<b>Change-orientation</b>	Does not address the importance of innovation; Does not address the benefits that could be inherent to the new process; Does not address the opportunities inherent to mutual support within the team		Mentions the importance of innovation; States the benefits that could be adherent to the new work process; If necessary, addresses the opportunities for mutual support within the team			Emphasizes the need for innovation; Explains with great enthusiasm the advantages of the new process for the organization; Provides a framework in which employees can support each other in implementing the process

## Example 2 of a past-behavior interview question

“Think of a situation in which it was very important that employees performed well for the success of your organization, but you had the impression that certain employees were dissatisfied, and your team did not work as efficiently or as quickly as you expected. Please describe both the situation and your exact response in detail.”

Notes:

Please rate the extent of the respective leadership behavior construct by ticking the appropriate number:

	1	2	3	4	5	
<b>Relations-orientation</b>	Shows no interest in employees' needs or feelings; Is not considerate of them; Does not try to involve employees in designing new work processes; Offers no support		Acknowledges employees' needs and feelings; Tries to be considerate; Allows employees to be involved in designing new work processes if so desired; Offers individual support if necessary		Asks employees about their needs and feelings; Is very considerate; Involves employees in designing work processes; Offers comprehensive support	
<b>Task-orientation</b>	Does not explain the priorities, tasks or responsibilities; States no expectations; Does not monitor work performance		Informs employees about priorities, tasks and responsibilities; Addresses what is expected of the employees; Checks work performance sometimes		Explains priorities, tasks and responsibilities in detail; Clearly communicates what is expected of the employees; Analyzes the problem; Monitors work performance consistently	
<b>Change-orientation</b>	Does not address the possibilities of mutual support; Does not address the opportunities that could be inherent to the challenging situation; Shows no interest in creative proposals		Addresses the possibilities of mutual support within the team where applicable; Mentions the opportunities that could be inherent to the challenging situation; Is open to creative solution proposals from the team		Provides a framework in which employees can mutually support each other; Explains with much enthusiasm the opportunities of the challenging situation; Encourages employees to search for creative solutions	

Example 1 of a future-behavior interview question

“Imagine that the team that you manage has been given the responsibility for a novel project that is very important for your organization. Your employees are already working at full capacity, and you must tell them about the project, but you do not yet have a clear picture of the project’s likelihood of success. Please describe how exactly you would proceed in this situation.”

Notes:

Please rate the extent of the respective leadership behavior construct by ticking the appropriate number:

<b>Relations-orientation</b>	<b>1</b> Does not acknowledge employees’ opinions, perceptions or feelings; Is not considerate; Does not involve employees in planning; Does not provide support	<b>2</b>	<b>3</b> Acknowledges employees’ opinions, perceptions and feelings and reacts to them if necessary, Is considerate; Involves employees in planning if so desired; Offers support	<b>4</b>	<b>5</b> Asks employees about opinions, perceptions and feelings and acknowledges these, Is very considerate of others; Involves employees in planning; Provides comprehensive support
<b>Task-orientation</b>	<b>1</b> Does not develop a plan; Does not explain tasks or responsibilities; States no expectations; Does not give specific instructions	<b>2</b>	<b>3</b> Develops ideas for the implementation of the project; Informs employees about tasks and responsibilities; Discusses what is expected of them; Gives instructions	<b>4</b>	<b>5</b> Develops a detailed implementation plan; Explains tasks and responsibilities; Analyzes the situation; Clearly communicates what is expected of the employees; Gives feasible instructions
<b>Change-orientation</b>	<b>1</b> Does not mention the importance of flexibility and innovation; Does not explain the advantages and opportunities that may be inherent to the new project; Does not address what the team could achieve	<b>2</b>	<b>3</b> Mentions the importance of flexibility and innovation; Explains the advantages and opportunities for the organization that could be inherent to the new project; Paints a picture of what the team could achieve	<b>4</b>	<b>5</b> Emphasizes the need for flexibility and innovation; Explains with great enthusiasm the advantages and opportunities for the organization that are inherent in the new project; Paints a clear picture of what the team can achieve



## Example 2 of a future-behavior interview question

“Imagine you vowed to make a case for hiring an additional employee to reduce your team’s workload. Due to the general poor economic situation, the necessary financial funds were not granted and you are unable to deliver on your promise to the team. Because of a new assignment that your team has received, the workload is likely to increase even more. Please describe how exactly you would proceed in this situation.”

Notes:

Please rate the extent of the respective leadership behavior construct by ticking the appropriate number:

<b>Relations-orientation</b>	<b>1</b> Shows no interest in employees' situation or feelings; Does not provide support; Does not consult with employees	<b>2</b>	<b>3</b> Acknowledges employees' situation and feelings; Provides support if necessary; Involves employees in problem-solving if so desired	<b>4</b>	<b>5</b> Acknowledges employees' situation and feelings; Provides comprehensive support; Involves employees in problem-solving; Asks employees which workload is feasible
<b>Task-orientation</b>	<b>1</b> Does not explain what needs to be done; Does not assign responsibilities, States no expectations; Does not give specific instructions	<b>2</b>	<b>3</b> Develops a plan; Informs employees about tasks and responsibilities; Discusses what is expected of them; Gives instructions	<b>4</b>	<b>5</b> Develops a detailed plan to cover tasks as best as possible without further employees; Explains exactly what is expected of the employees; Analyzes the situation; Gives feasible instructions
<b>Change-orientation</b>	<b>1</b> Does not emphasize the importance of flexibility and maximum work-effort; Does not address the advantages and opportunities that may be inherent to the new assignment; Does not address opportunities for mutual support	<b>2</b>	<b>3</b> Emphasizes the importance of flexibility and maximum work-effort; Explains the benefits and opportunities for the organization that may be inherent to the new assignment; Considers the possibilities of mutual support within the team where applicable	<b>4</b>	<b>5</b> Emphasizes the need for flexibility and work-effort; Explains with optimism the opportunities for the organization which are inherent to the new assignment; Provides a framework in which employees support each other as a team

## References

- Ahn, J., Lee, S., & Yun, S. (2018). Leaders' core self-evaluation, ethical leadership, and employees' job performance: The moderating role of employees' exchange ideology. *Journal of Business Ethics*, 148, 457–470. <https://doi.org/10.1007/s10551-016-3030-0>.
- Allen, N. J., & Meyer, J. P. (1990). The measurement and antecedents of affective, continuance and normative commitment to the organization. *Journal of Occupational Psychology*, 63, 1–18. <https://doi.org/10.1111/j.2044-8325.1990.tb00506.x>.
- Antonakis, J., Bastardoz, N., Jacquart, P., & Shamir, B. (2016). Charisma: An ill-defined and ill-measured gift. *Annual Review of Organizational Psychology and Organizational Behavior*, 3, 293–319. <https://doi.org/10.1146/annurev-orgpsych-041015-062305>.
- Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010). On making causal claims: A review and recommendations. *The Leadership Quarterly*, 21, 1086–1120. <https://doi.org/10.1016/j.leaqua.2010.10.010>.
- Antonakis, J., & Dietz, J. (2011). Looking for validity or testing it? The perils of stepwise regression, extreme-scores analysis, heteroscedasticity, and measurement error. *Personality and Individual Differences*, 50, 409–415. <https://doi.org/10.1016/j.paid.2010.09.014>.
- Antonakis, J., & House, R. J. (2014). Instrumental leadership: Measurement and extension of transformational-transactional leadership theory. *The Leadership Quarterly*, 25, 746–771. <https://doi.org/10.1016/j.leaqua.2014.04.005>.
- Atkins, P. W. B., & Wood, R. E. (2002). Self- versus others' ratings as predictors of assessment center ratings: Validation evidence for 360-degree feedback programs. *Personnel Psychology*, 55, 871–904. <https://doi.org/10.1111/j.1744-6570.2002.tb00133.x>.
- Atwater, L. E., & Yammarino, F. J. (1992). Does self-other agreement on leadership perceptions moderate the validity of leadership and performance predictions? *Personnel Psychology*, 45, 141–164. <https://doi.org/10.1111/j.1744-6570.1992.tb00848.x>.
- Avolio, B. J., Bass, B. M., & Jung, D. I. (1999). Re-examining the components of transformational and transactional leadership using the Multifactor Leadership Questionnaire. *Journal of Occupational and Organizational Psychology*, 72, 441–462. <https://doi.org/10.1348/096317999166789>.
- Bartko, J. J. (1976). On various intraclass correlation reliability coefficients. *Psychological Bulletin*, 83, 762–765. <https://doi.org/10.1037/0033-2909.83.5.762>.
- Bentler, P. M., & Chou, C.-P. (1987). Practical issues in structural modeling. *Sociological Methods & Research*, 16, 78–117. <https://doi.org/10.1177/0049124187016001004>.
- Bernardin, H. J., & Buckley, M. R. (1981). Strategies in rater training. *Academy of Management Review*, 6, 205–212. <https://doi.org/10.5465/AMR.1981.4287782>.
- Bernecker, K., Herrmann, M., Brandstätter, V., & Job, V. (2017). Implicit theories about willpower predict subjective well-being. *Journal of Personality*, 85, 136–150. <https://doi.org/10.1111/jopy.12225>.
- Bernerth, J. B., Cole, M. S., Taylor, E. C., & Walker, H. J. (2018). Control variables in leadership research: A qualitative and quantitative review. *Journal of Management*, 44, 131–160. <https://doi.org/10.1177/0149206317690586>.
- Bliese, P. D. (2000). Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis. In K. J. Klein, & S. W. J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 349–381). San Francisco, CA, US: Jossey-Bass.
- Bollen, K. A. (1989). *Structural equations with latent variables*. Oxford: John Wiley & Sons.
- Borgmann, L., Rowold, J., & Bormann, K. C. (2016). Integrating leadership research: A meta-analytical test of Yukl's meta-categories of leadership. *Personnel Review*, 45, 1340–1366. <https://doi.org/10.1108/PR-07-2014-0145>.
- Brown, M. E., Treviño, L. K., & Harrison, D. A. (2005). Ethical leadership: A social learning perspective for construct development and testing. *Organizational Behavior and Human Decision Processes*, 97, 117–134. <https://doi.org/10.1016/j.obhdp.2005.03.002>.
- Ceri-Booms, M., Curşeu, P. L., & Oerlemans, L. A. G. (2017). Task and person-focused leadership behaviors and team performance: A meta-analysis. *Human Resource Management Review*, 27, 178–192. <https://doi.org/10.1016/j.hrmr.2016.09.010>.
- Conger, J. A., & Kanungo, R. N. (1998). *Charismatic leadership in organizations*. Thousand Oaks, CA: Sage Publications, Inc.
- Cropanzano, R., & Mitchell, M. S. (2005). Social exchange theory: An interdisciplinary review. *Journal of Management*, 31, 874–900. <https://doi.org/10.1177/0149206305279602>.
- DeRue, D. S., Nahrgang, J. D., Wellman, N., & Humphrey, S. E. (2011). Trait and behavioral theories of leadership: An integration and meta-analytic test of their relative validity. *Personnel Psychology*, 64, 7–52. <https://doi.org/10.1111/j.1744-6570.2010.01201.x>.
- Fan, J., Stuhlmán, M., Chen, L., & Weng, Q. (2016). Both general domain knowledge and situation assessment are needed to better understand how SJTs work. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 9, 43–47. <https://doi.org/10.1017/iop.2015.114>.
- Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin*, 51, 327–358. <https://doi.org/10.1037/h0061470>.
- Fleener, J. W., Smither, J. W., Atwater, L. E., Braddy, P. W., & Sturm, R. E. (2010). Self-other rating agreement in leadership: A review. *The Leadership Quarterly*, 21, 1005–1034. <https://doi.org/10.1016/j.leaqua.2010.10.006>.
- Gentry, W. A., Hannum, K. M., Ekelund, B. Z., & de Jong, A. (2007). A study of the discrepancy between self- and observer-ratings on managerial derailment characteristics of European managers. *European Journal of Work and Organizational Psychology*, 16, 295–325. <https://doi.org/10.1080/13594320701394188>.
- Geyer, A. L. J., & Steyrer, J. M. (1998). Transformational leadership and objective performance in banks. *Applied Psychology: An International Review*, 47, 397–420. <https://doi.org/10.1080/0269994983777917>.
- Goffin, R. D., & Anderson, D. W. (2007). The self-rater's personality and self-other disagreement in multi-source performance ratings: Is disagreement healthy? *Journal of Managerial Psychology*, 22, 271–289. <https://doi.org/10.1108/02683940710733098>.
- Grömping, U. (2006). Relative importance for linear regression in R: The package relaimpo. *Journal of Statistical Software*, 17. <https://doi.org/10.18637/jss.v017.i01>.
- Hansbrough, T. K., Lord, R. G., & Schyns, B. (2015). Reconsidering the accuracy of follower leadership ratings. *The Leadership Quarterly*, 26, 220–237. <https://doi.org/10.1016/j.leaqua.2014.11.006>.
- Harms, P. D., Credé, M., Tynan, M., Leon, M., & Jeung, W. (2017). Leadership and stress: A meta-analytic review. *The Leadership Quarterly*, 28, 178–194. <https://doi.org/10.1016/j.leaqua.2016.10.006>.
- Hiller, N. J., DeChurch, L. A., Murase, T., & Doty, D. (2011). Searching for outcomes of leadership: A 25-year review. *Journal of Management*, 37, 1137–1177. <https://doi.org/10.1177/0149206310393520>.
- Hoffman, B. J., Woehr, D. J., Maldagen-Youngjohn, R., & Lyons, B. D. (2011). Great man or great myth? A quantitative review of the relationship between individual differences and leader effectiveness. *Journal of Occupational and Organizational Psychology*, 84, 347–381. <https://doi.org/10.1348/096317909x485207>.
- Huffcutt, A. I., Conway, J. M., Roth, P. L., & Stone, N. J. (2001). Identification and meta-analytic assessment of psychological constructs measured in employment interviews. *Journal of Applied Psychology*, 86, 897–913. <https://doi.org/10.1037/0021-9010.86.5.897>.
- Huffcutt, A. I., Weekley, J. A., Wiesner, W. H., DeGroot, T. G., & Jones, C. (2001). Comparison of situational and behavior description interview questions for higher-level positions. *Personnel Psychology*, 54, 619–644. <https://doi.org/10.1111/j.1744-6570.2001.tb00225.x>.
- Hunter, S. T., Bedell-Avers, K. E., & Mumford, M. D. (2007). The typical leadership study: Assumptions, implications, and potential remedies. *The Leadership Quarterly*, 18, 435–446. <https://doi.org/10.1016/j.leaqua.2007.07.001>.
- Ingold, P. V., Kleinmann, M., König, C. J., & Melchers, K. G. (2015). Shall we continue or stop disapproving of self-presentation? Evidence on impression management and faking in a selection context and their relation to job performance. *European Journal of Work and Organizational Psychology*, 24, 420–432. <https://doi.org/10.1080/1359432X.2014.915215>.
- Jackson, T. A., Meyer, J. P., & Wang, X.-H. (2013). Leadership, commitment, and culture: A meta-analysis. *Journal of Leadership & Organizational Studies*, 20, 84–106. <https://doi.org/10.1177/1548051812466919>.
- James, L. R., Demaree, R. G., & Wolf, G. (1984). Estimating within-group interrater reliability with and without response bias. *Journal of Applied Psychology*, 69, 85–98. <https://doi.org/10.1037/0021-9010.69.1.85>.
- James, L. R., Demaree, R. G., & Wolf, G. (1993).  $r_{wg}$ : An assessment of within-group interrater agreement. *Journal of Applied Psychology*, 78, 306–309. <https://doi.org/10.1037/0021-9010.78.2.306>.
- Jansen, A., Melchers, K. G., Lievens, F., Kleinmann, M., Brändli, M., Fraefel, L., & König, C. J. (2013). Situation assessment as an ignored factor in the behavioral consistency paradigm underlying the validity of personnel selection procedures. *Journal of Applied Psychology*, 98, 326–341. <https://doi.org/10.1037/a0031257>.
- Janz, T. (1982). Initial comparisons of patterned behavior description interviews versus unstructured interviews. *Journal of Applied Psychology*, 67, 577–580. <https://doi.org/10.1037/0021-9010.67.5.577>.
- Johnson, J. W. (2000). A heuristic method for estimating the relative weight of predictor variables in multiple regression. *Multivariate Behavioral Research*, 35, 1–19. <https://doi.org/10.1207/S15327906MBR3501.1>.
- Judge, T. A., Bono, J. E., Erez, A., & Locke, E. A. (2005). Core self-evaluations and job and life satisfaction: The role of self-concordance and goal attainment. *Journal of Applied Psychology*, 90, 257–268. <https://doi.org/10.1037/0021-9010.90.2.257>.
- Judge, T. A., & Kammeyer-Mueller, J. D. (2011). Implications of core self-evaluations for a changing organizational context. *Human Resource Management Review*, 21, 331–341. <https://doi.org/10.1016/j.hrmr.2010.10.003>.
- Judge, T. A., Klinger, R. L., & Simon, L. S. (2010). Time is on my side: Time, general mental ability, human capital, and extrinsic career success. *Journal of Applied Psychology*, 95, 92–107. <https://doi.org/10.1037/a0017594>.
- Judge, T. A., & Piccolo, R. F. (2004). Transformational and transactional leadership: A meta-analytic test of their relative validity. *Journal of Applied Psychology*, 89, 755–768. <https://doi.org/10.1037/0021-9010.89.5.755>.
- Junker, N. M., & van Dick, R. (2014). Implicit theories in organizational settings: A systematic review and research agenda of implicit leadership and followership theories. *The Leadership Quarterly*, 25, 1154–1173. <https://doi.org/10.1016/j.leaqua.2014.09.002>.
- Kauffeld, S., & Lehmann-Willenbrock, N. (2012). Meetings matter: Effects of team meetings on team and organizational success. *Small Group Research*, 43, 130–158. <https://doi.org/10.1177/1046496411429599>.
- Kauffeld, S., & Meyers, R. A. (2009). Complaint and solution-oriented circles: Interaction patterns in work group discussions. *European Journal of Work and Organizational Psychology*, 18, 267–294. <https://doi.org/10.1080/13594320701693209>.
- Klehe, U.-C., & Anderson, N. (2007). Working hard and working smart: Motivation and ability during typical and maximum performance. *Journal of Applied Psychology*, 92, 978–992. <https://doi.org/10.1037/0021-9010.92.4.978>.
- Klehe, U.-C., König, C. J., Richter, G. M., Kleinmann, M., & Melchers, K. G. (2008). Transparency in structured interviews: Consequences for construct and criterion-related validity. *Human Performance*, 21, 107–137. <https://doi.org/10.1080/08959280801917636>.
- Kleinmann, M., Ingold, P. V., Lievens, F., Jansen, A., Melchers, K. G., & König, C. J. (2011). A different look at why selection procedures work: The role of candidates' ability to identify criteria. *Organizational Psychology Review*, 1, 128–146. <https://doi.org/10.1080/0269994983777917>.

- org/10.1177/2041386610387000.
- Kozlowski, S. W. J. (2012). The nature of organizational psychology. In S. W. J. Kozlowski (Vol. Ed.), *The Oxford handbook of organizational psychology. Vol. 1. The Oxford handbook of organizational psychology* (pp. 3–21). New York, NY: Oxford University Press.
- Krajewski, H. T., Goffin, R. D., McCarthy, J. M., Rothstein, M. G., & Johnston, N. (2006). Comparing the validity of structured interviews for managerial-level employees: Should we look to the past or focus on the future? *Journal of Occupational and Organizational Psychology*, 79, 411–432. <https://doi.org/10.1348/096317905X68790>.
- Kurtessis, J. N., Eisenberger, R., Ford, M. T., Buffardi, L. C., Stewart, K. A., & Adis, C. S. (2017). Perceived organizational support: A meta-analytic evaluation of organizational support theory. *Journal of Management*, 43, 1854–1884. <https://doi.org/10.1177/0149206315575554>.
- Landis, R. S., Beal, D. J., & Tesluk, P. E. (2000). A comparison of approaches to forming composite measures in structural equation models. *Organizational Research Methods*, 3, 186–207. <https://doi.org/10.1177/109442810032003>.
- Latham, G. P., Saari, L. M., Pursell, E. D., & Campion, M. A. (1980). The situational interview. *Journal of Applied Psychology*, 65, 422–427. <https://doi.org/10.1037/0021-9010.65.4.422>.
- Lehmann-Willenbrock, N., & Allen, J. A. (2014). How fun are your meetings? Investigating the relationship between humor patterns in team interactions and team performance. *Journal of Applied Psychology*, 99, 1278–1287. <https://doi.org/10.1037/a0038083>.
- Liden, R. C., & Antonakis, J. (2009). Considering context in psychological leadership research. *Human Relations*, 62, 1587–1605. <https://doi.org/10.1177/0018726709346374>.
- Lievens, F., & Motowidlo, S. J. (2016). Situational judgment tests: From measures of situational judgment to measures of general domain knowledge. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 9, 3–22. <https://doi.org/10.1017/iop.2015.71>.
- Macan, T. (2009). The employment interview: A review of current studies and directions for future research. *Human Resource Management Review*, 19, 203–218. <https://doi.org/10.1016/j.hrmr.2009.03.006>.
- Mangold International (2010). INTERACT quick start manual V2.4. Available online at [mangold-international.com](http://mangold-international.com).
- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1, 30–46. <https://doi.org/10.1037/1082-989X.1.1.30>.
- Melchers, K. G., Klehe, U. C., Richter, G. M., Kleinmann, M., König, C. J., & Lievens, F. (2009). “I know what you want to know”: The impact of interviewees’ ability to identify criteria on interview performance and construct-related validity. *Human Performance*, 22, 355–374. <https://doi.org/10.1080/08959280903120295>.
- Melchers, K. G., & Kleinmann, M. (2016). Why situational judgment is a missing component in the theory of SJTs. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 9, 29–34. <https://doi.org/10.1017/iop.2015.111>.
- Meyer, J. P., & Allen, N. J. (1997). *Commitment in the workplace: Theory, research, and application*. Thousand Oaks, CA: Sage Publications, Inc.
- Meyer, J. P., Allen, N. J., & Smith, C. A. (1993). Commitment to organizations and occupations: Extension and test of a three-component conceptualization. *Journal of Applied Psychology*, 78, 538–551. <https://doi.org/10.1037/0021-9010.78.4.538>.
- Miao, C., Humphrey, R. H., & Qian, S. (2016). Leader emotional intelligence and subordinate job satisfaction: A meta-analysis of main, mediator, and moderator effects. *Personality and Individual Differences*, 102, 13–24. <https://doi.org/10.1016/j.paid.2016.06.056>.
- Montano, D., Reeske, A., Franke, F., & Hüffmeier, J. (2017). Leadership, followers’ mental health and job performance in organizations: A comprehensive meta-analysis from an occupational health perspective. *Journal of Organizational Behavior*, 38, 327–350. <https://doi.org/10.1002/job.2124>.
- Morgeson, F. P., Reid, M. H., & Campion, M. A. (2005). Selecting individuals in team settings: The importance of social skills, personality characteristics, and teamwork knowledge. *Personnel Psychology*, 58, 583–611. <https://doi.org/10.1111/j.1744-6570.2005.655.x>.
- Motowidlo, S. J., & Beier, M. E. (2010). Differentiating specific job knowledge from implicit trait policies in procedural knowledge measured by a situational judgment test. *Journal of Applied Psychology*, 95, 321–333. <https://doi.org/10.1037/a0017975>.
- Mumford, M. D., Zaccaro, S. J., Harding, F. D., Jacobs, T. O., & Fleishman, E. A. (2000). Leadership skills for a changing world: Solving complex social problems. *The Leadership Quarterly*, 11, 11–35. [https://doi.org/10.1016/S1048-9843\(99\)00041-7](https://doi.org/10.1016/S1048-9843(99)00041-7).
- Muthén, L. K., & Muthén, B. O. (2013). *Mplus user’s guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.
- Ng, T. W. H., Eby, L. T., Sorensen, K. L., & Feldman, D. C. (2005). Predictors of objective and subjective career success: A meta-analysis. *Personnel Psychology*, 58, 367–408. <https://doi.org/10.1111/j.1744-6570.2005.00515.x>.
- Nielsen, K., Nielsen, M. B., O’bonnaya, C., Känslä, M., Saari, E., & Isaksson, K. (2017). Workplace resources to improve both employee well-being and performance: A systematic review and meta-analysis. *Work & Stress*, 31, 101–120. <https://doi.org/10.1080/02678373.2017.1304463>.
- Paustian-Underdahl, S. C., Walker, L. S., & Woehr, D. J. (2014). Gender and perceptions of leadership effectiveness: A meta-analysis of contextual moderators. *Journal of Applied Psychology*, 99, 1129–1145. <https://doi.org/10.1037/a0036751>.
- Peus, C., Braun, S., & Frey, D. (2013). Situation-based measurement of the full range of leadership model - development and validation of a situational judgment test. *The Leadership Quarterly*, 24, 777–795. <https://doi.org/10.1016/j.leaqua.2013.07.006>.
- Piccolo, R. F., Bono, J. E., Heinritz, K., Rowold, J., Duehr, E., & Judge, T. A. (2012). The relative impact of complementary leader behaviors: Which matter most? *The Leadership Quarterly*, 23, 567–581. <https://doi.org/10.1016/j.leaqua.2011.12.008>.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88, 879–903. <https://doi.org/10.1037/0021-9010.88.5.879>.
- Podsakoff, P. M., MacKenzie, S. B., Moorman, R. H., & Fetter, R. (1990). Transformational leader behaviors and their effects on followers’ trust in leader, satisfaction, and organizational citizenship behaviors. *The Leadership Quarterly*, 1, 107–142. [https://doi.org/10.1016/1048-9843\(90\)90009-7](https://doi.org/10.1016/1048-9843(90)90009-7).
- Ree, M. J., & Carretta, T. R. (2006). The role of measurement error in familiar statistics. *Organizational Research Methods*, 9, 99–112. <https://doi.org/10.1177/1094428105283192>.
- Roch, S. G., Woehr, D. J., Mishra, V., & Kieszczynska, U. (2012). Rater training revisited: An updated meta-analytic review of frame-of-reference training. *Journal of Occupational and Organizational Psychology*, 85, 370–395. <https://doi.org/10.1111/j.2044-8325.2011.02045.x>.
- Rode, J. C., Arthaud-Day, M., Ramaswami, A., & Howes, S. (2017). A time-lagged study of emotional intelligence and salary. *Journal of Vocational Behavior*, 101, 77–89. <https://doi.org/10.1016/j.jvb.2017.05.001>.
- Rowold, J., & Borgmann, L. (2014). Interpersonal affect and the assessment of and interrelationship between leadership constructs. *Leadership*, 10, 308–325. <https://doi.org/10.1177/1742715013486046>.
- Sackett, P. R. (2007). Revisiting the origins of the typical-maximum performance distinction. *Human Performance*, 20, 179–185. <https://doi.org/10.1080/08959280701332968>.
- Sackett, P. R., Zedeck, S., & Fogli, L. (1988). Relations between measures of typical and maximum job performance. *Journal of Applied Psychology*, 73, 482–486. <https://doi.org/10.1037/0021-9010.73.3.482>.
- Sala, F. (2003). Executive blind spots: Discrepancies between self- and other-ratings. *Consulting Psychology Journal: Practice and Research*, 55, 222–229. <https://doi.org/10.1037/1061-4087.55.4.222>.
- Schermele-Heng, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research*, 8(2), 23–74.
- Schutte, N. S., Malouff, J. M., Hall, L. E., Haggerty, D. J., Cooper, J. T., Golden, C. J., & Dornheim, L. (1998). Development and validation of a measure of emotional intelligence. *Personality and Individual Differences*, 25, 167–177. [https://doi.org/10.1016/S0191-8869\(98\)00001-4](https://doi.org/10.1016/S0191-8869(98)00001-4).
- Schyns, B., & Schilling, J. (2013). How bad are the effects of bad leaders? A meta-analysis of destructive leadership and its outcomes. *The Leadership Quarterly*, 24, 138–158. <https://doi.org/10.1016/j.leaqua.2012.09.001>.
- Shaffer, J. A., DeGeest, D., & Li, A. (2016). Tackling the problem of construct proliferation: A guide to assessing the discriminant validity of conceptually related constructs. *Organizational Research Methods*, 19, 80–110. <https://doi.org/10.1177/1094428115598239>.
- Stogdill, R. M., & Coons, A. E. (1957). *Leader behavior: Its description and measurement*. Oxford, England: Ohio State University, Bureau of Busin.
- Stogdill, R. M., Goode, O. S., & Day, D. R. (1962). New leader behavior description subscales. *The Journal of Psychology: Interdisciplinary and Applied*, 54, 259–269. <https://doi.org/10.1080/00223980.1962.9713117>.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Allyn & Bacon/Pearson Education.
- Topp, C. W., Østergaard, S. D., Søndergaard, S., & Bech, P. (2015). The WHO-5 Well-Being Index: A systematic review of the literature. *Psychotherapy and Psychosomatics*, 84, 167–176. <https://doi.org/10.1159/000376585>.
- Van Iddekinge, C. H., Raymark, P. H., Eidson, C. E., Jr., & Attenweiler, W. J. (2004). What do structured selection interviews really measure? The construct validity of behavior description interviews. *Human Performance*, 17, 71–93. [https://doi.org/10.1207/S15327043HUP1701\\_4](https://doi.org/10.1207/S15327043HUP1701_4).
- Van Iddekinge, C. H., Raymark, P. H., & Roth, P. L. (2005). Assessing personality with a structured employment interview: Construct-related validity and susceptibility to response inflation. *Journal of Applied Psychology*, 90, 536–552. <https://doi.org/10.1037/0021-9010.90.3.536>.
- Van Knippenberg, B., & Van Knippenberg, D. (2005). Leader self-sacrifice and leadership effectiveness: The moderating role of leader prototypicality. *Journal of Applied Psychology*, 90, 25–37. <https://doi.org/10.1037/0021-9010.90.1.25>.
- Vroom, V. H., & Jago, A. G. (2007). The role of the situation in leadership. *American Psychologist*, 62, 17–24. <https://doi.org/10.1037/0003-066X.62.1.17>.
- Walter, F., & Scheibe, S. (2013). A literature review and emotion-based model of age and leadership: New directions for the trait approach. *The Leadership Quarterly*, 24, 882–901. <https://doi.org/10.1016/j.leaqua.2013.10.003>.
- Wang, G., Van Iddekinge, C. H., Zhang, L., & Bishoff, J. (2019). Meta-analytic and primary investigations of the role of followers in ratings of leadership behavior in organizations. *Journal of Applied Psychology*, 104, 70–106. <https://doi.org/10.1037/apl0000345>.
- WHO (2001). *Mental health: New understanding, new hope. The world health report 2001*. Geneva: World Health Organization.
- Wilderom, C. P. M., van den Berg, P. T., & Wiersma, U. J. (2012). A longitudinal study of the effects of charismatic leadership and organizational culture on objective and perceived corporate performance. *The Leadership Quarterly*, 23, 835–848. <https://doi.org/10.1016/j.leaqua.2012.04.002>.
- Woehr, D. J., Ligon, A. C., Schmidt, P. B., Loughry, M. L., & Ohland, M. W. (2015). Justifying aggregation with consensus-based constructs: A review and examination of cutoff values for common aggregation indices. *Organizational Research Methods*, 18, 704–737. <https://doi.org/10.1177/1094428115582090>.
- Yukl, G. (1999). An evaluation of conceptual weaknesses in transformational and

- charismatic leadership theories. *The Leadership Quarterly*, 10, 285–305. [https://doi.org/10.1016/S1048-9843\(99\)00013-2](https://doi.org/10.1016/S1048-9843(99)00013-2).
- Yukl, G. (2012). Effective leadership behavior: What we know and what questions need more attention. *The academy of management perspectives*. Vol. 26. *The academy of management perspectives* (pp. 66–85). . <https://doi.org/10.5465/amp.2012.0088>.
- Yukl, G. (2013). *Leadership in organizations* (8th ed.). Boston: Pearson.
- Yukl, G., Gordon, A., & Taber, T. (2002). A hierarchical taxonomy of leadership behavior: Integrating a half century of behavior research. *Journal of Leadership & Organizational Studies*, 9, 15–32. <https://doi.org/10.1177/107179190200900102>.
- Yukl, G., Mahsud, R., Hassan, S., & Prussia, G. E. (2013). An improved measure of ethical leadership. *Journal of Leadership & Organizational Studies*, 20, 38–48. <https://doi.org/10.1177/1548051811429352>.
- Yukl, G., O'Donnell, M., & Taber, T. (2009). Influence of leader behaviors on the leader-member exchange relationship. *Journal of Managerial Psychology*, 24, 289–299. <https://doi.org/10.1108/02683940910952697>.
- Zaccaro, S. J., Green, J. P., Dubrow, S., & Kolze, M. (2018). Leader individual differences, situational parameters, and leadership outcomes: A comprehensive review and integration. *The Leadership Quarterly*, 29, 2–43. <https://doi.org/10.1016/j.leaqua.2017.10.003>.